

**United States Military Academy
West Point, New York 10996**

Shaping the ROTC Cohort

**OPERATIONS RESEARCH CENTER OF EXCELLENCE
TECHNICAL REPORT No. DSE-TR-0505
DTIC #: ADA434913**

Lead Analyst
Major Wiley P. Rittenhouse, M.S.
Analyst, Operations Research Center

Senior Investigator
Lieutenant Colonel Michael J. Kwinn, Jr., Ph.D.
Director, Operations Research Center of Excellence

Approved by
Colonel Michael L. McGinnis, Ph.D.
Professor and Head, Department of Systems Engineering

April 2005

The Operations Research Center of Excellence is supported by the
Assistant secretary of the Army (Financial Management & Comptroller)

Distribution A: Approved for public release; distribution is unlimited.

SHAPING THE ROTC COHORT

Lead Analyst

Major Wiley P. Rittenhouse

Analyst, Operations Research Center

Senior Investigator

Lieutenant Colonel Michael J. Kwinn, Jr., Ph.D.

Director, Operations Research Center of Excellence

OPERATIONS RESEARCH CENTER OF EXCELLENCE

TECHNICAL REPORT No. DSE-TR-0505

DTIC #: ADA434913

Approved by

Colonel Michael L. McGinnis, Ph.D.

Professor and Head, Department of Systems Engineering

April 2005

The Operations Research Center of Excellence is supported by the Assistant Secretary of the Army
(Financial Management & Comptroller)

This Research was sponsored by the Center for Accessions Research, US Army Accessions Command, Ft Knox, KY

Distribution A: Approved for public release; distribution is unlimited.

Abstract

U.S. Army Cadet Command (USACC) has developed a new commissioning model – Shaping the Cohort (STC) - to meet the future needs of the Army for commissioned officers. It is designed to shape each cohort to meet the Army’s specific needs in terms of component, academic disciplines, race/ethnic makeup goals, gender, and targeted missions. STC does this by determining and examining the “prime market” at a university and basing the detachment’s mission on penetration of that market as opposed to one based on past performance. It is believed that the STC model improves the method of determining missions.

To determine market potential, USACC conducted two surveys that included 62 colleges and universities and over 7600 students. The goals of the survey were to determine knowledge and perception of Army ROTC among students, segmentation of local markets, how the school markets differed, and the characteristics that could lead to participation in Army ROTC. The data gathered is used to determine how many students at each school fit the criteria for the prime market.

The Operations Research Center of Excellence (ORCEN) at West Point has provided an independent assessment of the model’s adequacy and to determine if and how it can be improved. We begin this report with a needs analysis to determine the real problem the model attempts to address. We then present our analysis of the STC model and the process for determining missions. In the next section we identify recommendations to improve the model and conclude with suggested additional tools for conducting this analysis.

About the Author(s)

MAJOR WILEY P. RITTENHOUSE is an instructor in the Department of Mathematics serving as an analyst in the Operations Research Center of Excellence at the United States Military Academy (USMA), West Point, New York. He received a BS in Mathematics from Kansas State University in 1994 and a MS in Statistics from Tulane University in 2003. His research interests include the application of statistical analyses and simulation to operations research. His email address is Wiley.Rittenhouse@usma.edu.

LIEUTENANT COLONEL MICHAEL J. KWINN, JR., is an Associate Professor in the Department of Systems Engineering and Director of the Operations Research Center of Excellence at the United States Military Academy, West Point. He has a B.S. Degree from USMA, M.S. degree in Systems and Industrial Engineering from the University of Arizona and a Ph.D. in Management Science and Information Systems from the University of Texas, Austin. His research interests include operational assessment methodology, efficiency analysis, recruiting analysis especially marketing effects and capability analysis and modeling. Lieutenant Colonel Kwinn may be contacted at Michael.Kwinn@usma.edu.

Acknowledgements

Thanks to Dr. Bobbie Foote, LTC Jeff Schamburg, MAJ Steve Henderson, and MAJ Grant Martin for their suggestions and advice during the course of this study.

Table of Contents

Abstract	iii
About the Author(s)	iv
Acknowledgements	iv
List of Figures	vii
List of Equations	viii
Chapter 1: Introduction.	1
1.1. A Brief History of the Army Reserve Officer Training Corps (ROTC).	1
1.2. Background.	1
1.3. The Purpose of STC.	2
1.3.1. The On-Campus Market Potential Study (OCMPS)	2
1.3.2. The Officer Accessions Strategy (OAS)	2
1.4. Approach.	3
1.5. Report Overview.	3
Chapter 2: STC Model and Current Implementation.	4
2.1. Overall Concept.	4
2.2. Implementation.	4
Chapter 3: Problems with Current Implementation.	6
3.1. Required Market Penetration in the “Working Model” Worksheet.	6
3.2. Underestimation of the Adjusted Prime Market.	7
3.3. Inconsistent transition from DST back to School Data worksheet.	8
3.4. Incorrect OCMPS values.	8
3.5. Duplicate Data.	9
Chapter 4: Recommendations.	10
4.1. Overview of recommended changes.	10
4.1.1. School Data Module (SDM).	10
4.1.2. OCMPS Data Module (ODM).	10
4.1.3. Mission Data Processing Module (MDPM).	11
4.1.4. Data Summary Module (SUM).	11
4.1.5. Historical Data Module (HIST).	11
4.2. A Detailed Look at the Changes.	11
4.2.1. Structure of the SDM.	12
4.2.2. Structure of the ODM.	13
4.2.2.1. Getting School Type.	13
4.2.2.2. Getting OCMPS Percentages.	14
4.2.3. Structure of the MDPM.	16
4.2.3.1. Getting the Recruitable Market Estimate.	16
4.2.3.2. Getting the Prime Market Estimate.	18
4.2.3.3. Getting the Commissioning Mission.	19
4.2.4. Suggested Functions for SUM.	19
4.3. A Word about Inputs.	20
4.4. Additional Modifications.	21
4.4.1. Delta to Track Mission Changes.	21
4.4.2. Alternate Minimum Mission Threshold Derived from Production History.	22
4.4.2.1. Initial Observations.	22
4.4.2.2. Methodology.	23

4.4.2.3.	An Important Note on Assumptions	25
4.4.2.4.	Calculation of the Percentile.....	26
4.4.2.4.1.	Directly from Historical Data.	26
4.4.2.4.2.	From the Approximate Normal Distribution.	26
4.4.2.5.	Implementation.	26
4.4.3.	Comparing Mission Numbers to the Estimated Production Distributions.....	28
Chapter 5:	Performance Measures.....	30
5.1.	Measuring Efficiency with Data Envelopment Analysis (DEA).	30
5.2.	The Purpose of This Chapter.....	30
5.3.	An Example Using Frontier Analyst Professional.	30
5.3.1.	Getting Data into Frontier Analyst.	31
5.3.2.	Setting Inputs and Outputs.....	32
5.3.3.	Analysis and Results Using the CCR Model.	34
5.3.3.1.	Example School 1: Appalachian State University.	36
5.3.3.2.	Example School 2: University of Portland.	39
5.3.4.	Analysis and Results Using the BCC Model.	41
5.3.5.	A Broader View at Future Application.....	43
Appendix A:	An Explanation of the Ryan-Joiner Test for Normality.....	46
Appendix B:	An Detailed Explanation of the Simulation Used to Determine the Alternate Minimum Mission Threshold.....	47
Appendix C:	Table of Abbreviations	53
Bibliography	54

List of Figures

Figure 1: STC Spreadsheets.....	4
Figure 2: STC Model, Overall Commission Mission	5
Figure 3: Required Market Penetration calculation in "Working Model".	6
Figure 4: Constant Value for Adjusted Prime Market.....	8
Figure 5: Function for Adjusted Prime Market.....	8
Figure 6: Proposed Model Implementation	12
Figure 7: OCMPS Data Module	13
Figure 8: Getting School Type from the tables in the ODM.	14
Figure 9: Getting OCMPS Percentages from the ODM	15
Figure 10: An example MDPM. Note that the school name is linked over from the SDM.	16
Figure 11: Calculating Adjustment 1.....	17
Figure 12: Calculating Adjustment 2.....	17
Figure 13: Calculating Adjustment 3. This result is the Recruitable Market Estimate	18
Figure 14: Calculating the Prime Market Estimate.....	18
Figure 15: Calculating the Commissioning Mission.	19
Figure 16: An example of a school type summary.	20
Figure 17: Using a delta to adjust "Mission" values.....	21
Figure 18: Histogram of sample correlation coefficients from the Ryan-Joiner test.....	23
Figure 19: Simulating school production from the whole distribution.....	24
Figure 20: A lower threshold of 6 is imposed.....	24
Figure 21: The lower threshold is set at the 34th percentile.	25
Figure 22: Examples of percentile calculation.....	27
Figure 23: Using the Alternate Minimum Mission Threshold.....	28
Figure 24: Data Input Dialog	31
Figure 25: STC data in the Data Viewer.....	32
Figure 26: Identifying the output variable.	33
Figure 27: Removing an individual school from the dataset.	33
Figure 28: The DEA Options dialog.....	34
Figure 29: The Efficiency Scores table.....	35
Figure 30: The Scores Distribution chart.....	35
Figure 31: The frontier plot.....	36
Figure 32: Comparison of Appalachian State to its reference set on the frontier plot.....	37
Figure 33: Potential improvements report (graph and table) for Appalachian State.	37
Figure 34: By-variable comparison to a reference set DMU.....	38
Figure 35: Comparison of the University of Portland to its reference set on the frontier plot.	39
Figure 36: The effect of implementing the potential improvements.	40
Figure 37: By-variable comparison of the University of Portland to Central Missouri State University.....	40
Figure 38: By-variable comparison of the University of Portland to the University of Hawaii at Manoa.	41
Figure 39: Changing to varying returns to scale (BCC model) in the DEA options dialog.	42
Figure 40: Efficiency Scores table from the BCC model.	42
Figure 41: Distribution of Scores chart from the BCC model.	43
Figure 42: An x-y plot of two input variables.....	44

List of Equations

Equation 1: Result of current calculation for Required Market Penetration by school type.....	7
Equation 2: Equation for expressing delta as a new required market penetration percentage.	22
Equation 3: Mathematical description of the efficient frontier.....	38

Chapter 1: Introduction

1.1. A Brief History of the Army Reserve Officer Training Corps (ROTC).

Army ROTC is the single largest commissioning source for second lieutenants in the active Army, US Army Reserve, and the Army National Guard. Between 60 and 70 percent of these second lieutenants are graduates of the ROTC program. ROTC officially began on 3 June 1916 when President Woodrow Wilson signed the National Defense Act of 1916. The first group of 133 cadets received their commissions during the 1919-1920 academic year. Since then, over half a million cadets have been commissioned as second lieutenants at one of the over 270 college and university ROTC programs. In 1986, the US Army Cadet Command (USACC) was established as a major subordinate command of the US Army Training & Doctrine Command (TRADOC). In 1998, cuts mandated by TRADOC resulted in a major reorganization of USACC and its three region headquarters. This began a trend in USACC toward trying to find ways to do “more with less”, of which the Shaping the Cohort (STC) model is a product (“Brief History of Army ROTC”).

1.2. Background.

The Reserve Officer Training Corps, or ROTC, at college campuses throughout the country has undergone significant changes in the past few years following the publication of the guidance presented in “College ROTC: The Way Ahead” (April, 2001, USACC). This policy document moved college-level ROTC programs away from an attrition-based model to a developmental leadership model. Though this is a significant departure for the management of the cadets in the ROTC programs, it did not immediately impact the process of recruiting the cadets into the program initially.

The US Army Cadet Command (USACC) affects the recruiting and retention of cadets in the national ROTC program most significantly through the allocation of recruiting goals and the allocation of resources. Until recently, the USACC determined these commissioning goals, or missions almost exclusively on a school’s historical production rates. USACC determined resource allocation similarly – in other words, resources tend to be allocated to places where they were used in the past, without considering any other factors. They soon realized that the shift in program styles required a change in missioning philosophy.

1.3. The Purpose of STC.

Cadet Command recognized that determining a school commissioning mission based solely on its historical production was keeping schools from penetrating into larger potential markets that might exist on campus. The Shaping the Cohort (STC) model paved the way for a move toward a market analysis approach, where market size was measured by determining the number of eligible students that possessed the desired characteristics and a likely interest in ROTC participation. Also, STC builds each ROTC cohort to meet certain targeted goals with regard to gender, race/ethnicity, and academic discipline mix (ADM). Two internal Cadet Command studies made this possible: the On-Campus Market Potential Study (OCMPS) and the Officer Accessions Strategy (OAS).

1.3.1. The On-Campus Market Potential Study (OCMPS).

The OCMPS was conducted in April of 2001 and 2002, and surveyed over 7600 students at 62 universities. The purpose of the study was to identify students who had the characteristics desired for future officers (identified as Scholar-Athlete-Leader, or SAL) as well as characteristics that made them more likely to participate in ROTC if they were exposed to the program (primarily identified as “First-Stringers”¹). Examination of the data in this study showed that the two strongest predictors of the values for SALs and First Stringers were the size of the school and its academic reputation. The aggregate data is represented by fourteen clusters based on these predictors. Work on this study is continuing, and is intended to eventually encompass all 272 schools.

1.3.2. The Officer Accessions Strategy (OAS).

The other main component of STC is to directly tailor the shape of each cohort with regard to race/ethnicity, gender, and ADM. The OAS determines the distribution of graduates at colleges in terms of these factors. This information is then used to help determine which portions of the cohort an individual program should be responsible for producing, which is passed on to the ROTC detachment as targeted missions (in addition to the overall production mission).

¹ These numbers are expressed as percentages of the school population; the SAL percentage is dependent on the First Stringers (i.e percentage of First Stringers who are also SALs).

1.4. Approach.

LTC William Warner (Division Chief, Operations Analysis Division, RROD, USACC) briefed us on the background and concept of STC (Warner, 2004), and provided a copy of the relevant computer files. We then began our examination of the model and its implementation. An interview was conducted with COL Michael Hoff (Director, Recruiting & Retention Operations Directorate, USACC) and LTC Warner to get their perspective on the role of the current model and the direction they thought it should take to meet the needs of USACC. During this interview, they stated that MG Alan W. Thrasher (CG, USACC) wants his subordinate commanders to have the freedom to exercise their professional military judgment with regard to the final commissioning mission assignments (Hoff, 2004). The implication of this is that the model outputs are intended as a starting point, which can be negotiated within the chain of command to a different value if sufficient justification is given. Additionally, they expressed a desire to improve the quality of their model inputs, and to try to find a measure of propensity (a desire and commitment to join ROTC) (Hoff, 2004). We then conducted a detailed analysis of the model and its implementation, along with an analysis of functions and data stored in the computer files. We conducted a follow-on interview with LTC Warner (Warner and others, 2005) to get clarification on certain aspects of the model and its implementation. Finally, we completed our analysis and developed our recommendations.

1.5. Report Overview.

In Chapter 2 of this report, we begin with a discussion of the overall concept of STC and how the commissioning mission model is implemented. In Chapter 3 we provide details about problems we saw with the current implementation. In Chapter 4, we will suggest a simplified structure for model implementation in Microsoft Excel, followed by detailed instructions on how to use Excel's built-in functions to handle the various calculations and data referencing. In this chapter we will also provide suggestions on some other modifications and additions that can be made, with an emphasis on an alternative minimum mission value. In Chapter 5, we will discuss the use of Data Envelopment Analysis to measure efficiency, and what data requirements may be needed to utilize this tool. Our conclusion is given in Chapter 6.

Chapter 2: STC Model and Current Implementation.

2.1. Overall Concept.

An initial recruitable market size is estimated using data on the number of baccalaureate degrees conferred using current data from the National Center for Education Statistics (NCES). That number is then adjusted by subtracting ineligible students (i.e. non-U.S. citizens, part-time students, or students older than 27 years of age) and adding students that join the ROTC program through campus partnerships². A final 6% reduction is applied if sister service ROTC programs are present at the same institution. The result of these adjustments is called the Adjusted Recruitable Market. The Adjust Prime Market (the market of interest) results from the product of the Adjusted Recruitable Market, the percentage of First Stringers, and the percentage of SALs. The Adjusted Mission is the product of the Adjusted Prime Market and a market penetration factor (expressed as a percentage), plus any applicable partnership cadets (see footnote 2). The Adjusted Mission represents the overall commission mission for a particular college or university.

2.2. Implementation.

STC is implemented in two separate Microsoft Excel spreadsheets. The bulk of the information on a particular school (including its number of baccalaureate-level graduates) and the relevant numbers from the OCMPS are stored in one file, while the process for adjusting the recruitable market size is carried out in a separate spreadsheet.

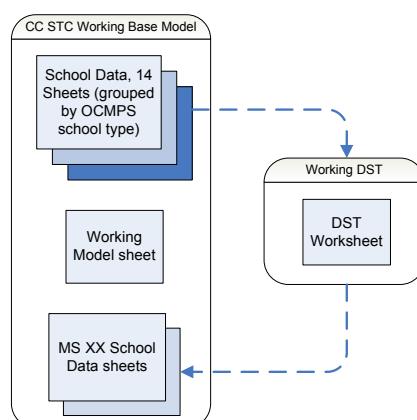


Figure 1: STC Spreadsheets

² If the partnership is with a school less than 45 minutes away and the distant school contributes at least 25% of a program's cadets, a 25% increase is applied to the Adjusted Recruitable Market. If a program does not meet both criteria, any cadets resulting from a partnership are added directly into the Adjusted Mission.

The arrows in the diagram in Figure 1 do not represent any actual transfer of data between the two files (there are no links for this action built into either file). It only indicates that data is transcribed by the user from one file to another. The “Working Model” worksheet is used to organize the data from the OCMPS and to ‘wargame’ values for commissioning mission and market penetration percentage based on school type. Figure 2 shows the entire implementation starting from the Recruitable Market (data from NCES) and ending with the Adjusted Mission.

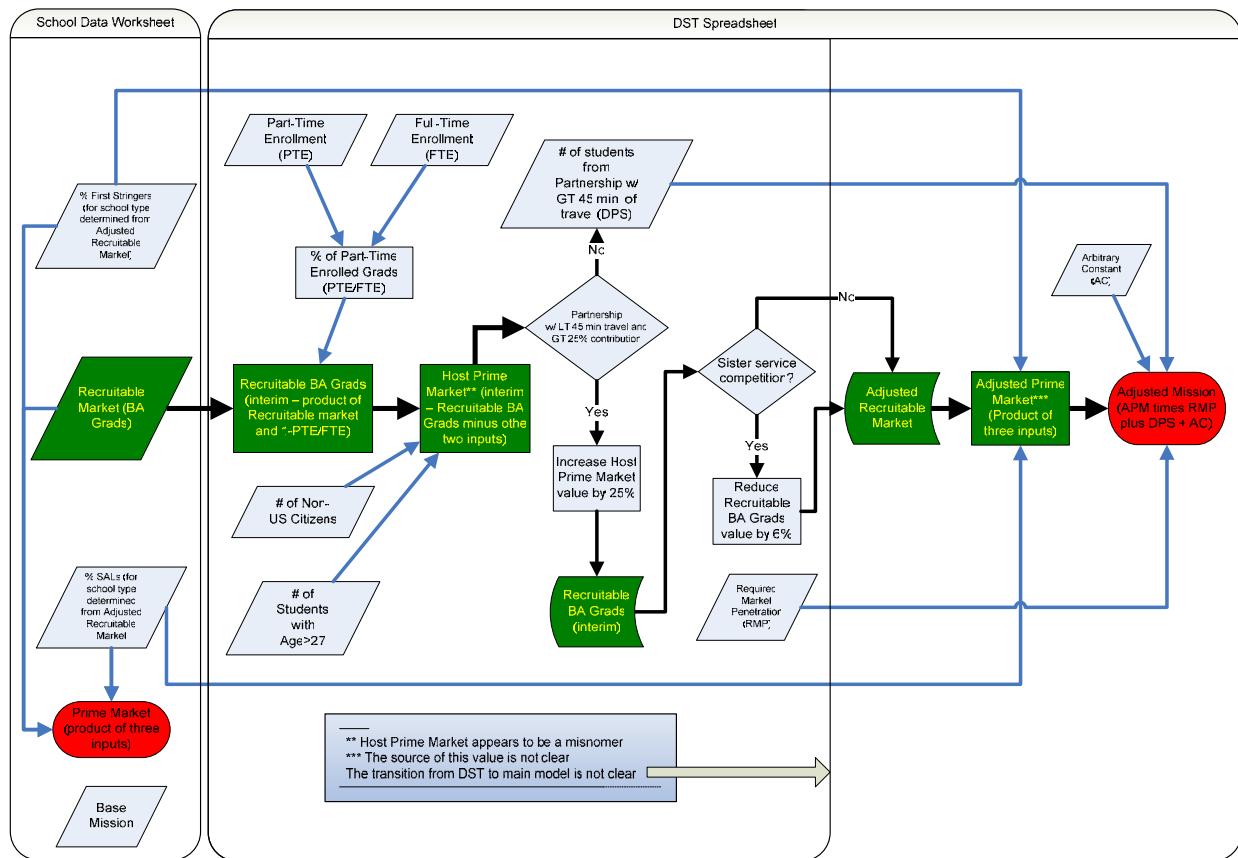


Figure 2: STC Model, Overall Commission Mission

Chapter 3: Problems with Current Implementation.

In general, very little actual calculation takes place within the model as implemented in the spreadsheets. Much of the data appears to be transferred manually (using a copy-and-paste operation) as opposed to direct cell addressing. The only calculation leading up to the figure for Adjusted Mission that is done consistently for all 272 schools is the Prime Market (shown in red on the left side of Figure 2). However, this adds no value to the outcome since no other process or calculation in the spreadsheet depends upon this figure (the prime market calculation only becomes meaningful when it is calculated from the Adjusted Recruitable Market). A column for Base Mission appears in the School Data worksheets, but has been transcribed manually from the Working Model worksheet (and actually does not appear to reflect the values for Base Mission calculated in the “Working Model” worksheet), and is also not used anywhere else in the spreadsheet. Other problems in the current implementation are discussed below.

3.1. Required Market Penetration in the “Working Model” Worksheet.

The left-hand side of this table is used to determine a base value for required market penetration and a basic mission value for each school type (see Figure 3). The “Basic Mission by School

BASE MODEL						
School Type	Prime Market	Avg Prime Market per School	Percent of Total Prime Market	# Schools in Category	Aggregate School Type Prod	Basic mission by School Type
Mega-High	6222	415	19.1%	15	873	58
Mega-Prestigious	1918	384	5.9%	5	269	54
Large-High	4432	164	13.8%	27	622	23
Large-Moderate	5430	118	16.6%	46	762	17
Large-Prestigious	932	155	2.9%	6	131	22
Large-Standard	1777	99	5.4%	18	249	14
Medium-High	823	75	2.5%	11	115	10
Medium-Moderate	2299	77	7.0%	30	322	11
Medium-Prestigious	1376	92	4.2%	15	193	13
Medium-Standard	1486	68	4.6%	22	208	9
Small-High	773	45	2.4%	17	108	6
Small-Moderate	348	32	1.1%	11	43	4
Small-Prestigious	99	33	0.3%	3	14	5
Small-Standard	403	34	1.3%	12	57	5
HBCUs	797	35	2.4%	23	112	5
Senior Military College	2298	383	7.0%	6	322	54
Military Junior College	1200	240	3.7%	5	168	34
Totals	32619		100%	272	4575	14.0%

Figure 3: Required Market Penetration calculation in "Working Model".

Type” provides a useful starting point for missions based on the OCMPS data³. These values are adjusted in the “Wargaming” section to account for known outlier school types. However, the “Required Market Penetration” values are not actually based on the data stored in each row, but merely reflect the ratio of Commission Mission (4575) to the total Prime Market size (32619), so that $4575/32619=0.14$ (these values are found at the bottom of the “Aggregate School Type Prod” and “Prime Market” columns, respectively). This is because the value for Prime Market used in the by-row calculations cancels out (see Equation 1).

$$\begin{aligned}
 \text{Required Mkt Penetration} &= \frac{\text{AggregateSchoolType Prod}}{\text{Prime Market}} \\
 &= \frac{(\text{Cmsn Msn})(\text{Percent of Total Prime Market})}{\text{Prime Market}} \\
 &= \frac{(\text{Cmsn Msn})\left(\frac{\text{Prime Market}}{\text{Prime Market Total}}\right)}{\text{Prime Market}} \\
 &= \frac{\text{Cmsn Msn}}{\text{Prime Market Total}} = \frac{4575}{32619} = 0.14
 \end{aligned}$$

Equation 1: Result of current calculation for Required Market Penetration by school type.

There is no real error in doing the calculation for Required Market Penetration in this manner – we merely point out that the structure of the worksheet obscures the source of the value, and does not really reflect the fact that it is merely calculating a value for required market penetration that is aggregated across the entire body of schools.

3.2. Underestimation of the Adjusted Prime Market.

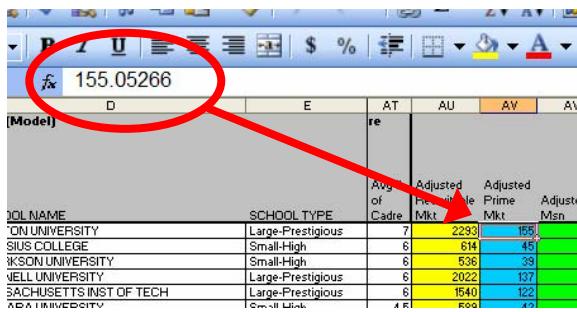
There are three sources for this underestimation. The first source is an accepted limitation of the model – that of only accounting for the percentage of First Stringers, and excluding other acceptable groups (i.e. Second Stringers). The second source is not accounting for the possibility that a student might fall into more than one category among the filters (students who are not U.S. citizens, are part-time, or are older than 27). It is possible, however, that this overlap is negligible. The last is determining the number of part-time students by calculating the ratio of part-time to full-time students (this calculation is performed in the Working DST spreadsheet). These two groups are mutually exclusive, and the resulting ratio

³ However, it should be noted that neither the Base Mission or Adjusted Mission values that appear in the “Working Model” worksheet appear to transfer to the “School Data” worksheets intact – new values appear in the respective columns on the “School Data” worksheets, implying further adjustment.

underestimates the proportion of full-time students. Dividing the number of full-time students by the school's total enrollment gives the proper proportion, which can be used in determining the estimated size of the prime market (this is shown in Figure 6 and discussed in more detail in section 4.2.3).

3.3. Inconsistent transition from DST back to School Data worksheet.

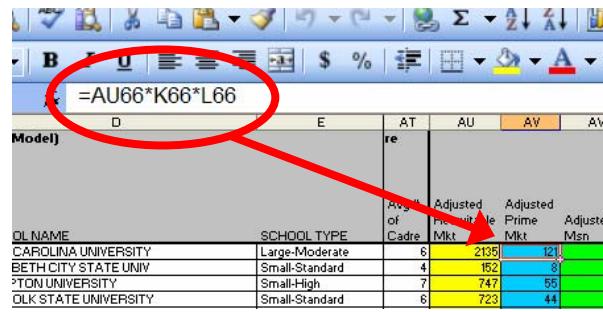
The first really significant input on the School Data worksheets is the Adjusted Recruitable Market, which results from the Working DST calculations. The Adjusted Prime Market is calculated from it using the OCMPS percentages. This is where the transition inconsistency exists: it is unclear where the calculation is performed. The Working DST does not have the same complete list of all 272 schools – it appears to be more of a tool that is intended to handle one input at a time and only produces one output. In the cases where the Adjusted Prime Market figure is entered as a constant in the School Data worksheets (see Figure 4), it is assumed that the number was produced using the Working DST. However, 26 schools have their Adjusted Prime Market value calculated in the School Data worksheet (see Figure 5), using values entered in the worksheet itself.



A screenshot of a Microsoft Excel spreadsheet. The top cell (B1) contains the value "155.05266". A red circle highlights this cell. Below it is a table with columns for ID NAME, SCHOOL TYPE, Avg. of Cadre, Avg. of Recruitable Mkt, Adjusted Prime Mkt, and Adjusted Msn. The table shows data for various schools like CAROLINA UNIVERSITY, BETH CITY STATE UNIV, STON UNIVERSITY, and OLK STATE UNIVERSITY.

ID NAME	SCHOOL TYPE	Avg. of Cadre	Avg. of Recruitable Mkt	Adjusted Prime Mkt	Adjusted Msn
CAROLINA UNIVERSITY	Large-Prestigious	7	2293	155	
SIUS COLLEGE	Small-High	6	614	45	
NSON UNIVERSITY	Small-High	6	536	39	
JELL UNIVERSITY	Large-Prestigious	6	2022	137	
SACHUSETTS INST OF TECH	Large-Prestigious	6	1540	122	
ADAIR UNIVERSITY	Small-High	6	609	42	

Figure 4: Constant Value for Adjusted Prime Market



A screenshot of a Microsoft Excel spreadsheet. The top cell (B1) contains the formula "=AU66*K66*L66". A red circle highlights this cell. Below it is a table with columns for ID NAME, SCHOOL TYPE, Avg. of Cadre, Avg. of Recruitable Mkt, Adjusted Prime Mkt, and Adjusted Msn. The table shows data for various schools like CAROLINA UNIVERSITY, BETH CITY STATE UNIV, STON UNIVERSITY, and OLK STATE UNIVERSITY.

ID NAME	SCHOOL TYPE	Avg. of Cadre	Avg. of Recruitable Mkt	Adjusted Prime Mkt	Adjusted Msn
CAROLINA UNIVERSITY	Large-Moderate	6	2135	121	
BETH CITY STATE UNIV	Small-Standard	4	152	8	
STON UNIVERSITY	Small-High	7	747	55	
OLK STATE UNIVERSITY	Small-Standard	6	723	44	

Figure 5: Function for Adjusted Prime Market

3.4. Incorrect OCMPS values.

In 13 cases, the wrong OCMPS percentages are entered for that school. Closer examination revealed that the percentages entered were based on the school size that resulted from the 'BA Grads' value (see Figure 2) value instead of the Adjusted Recruitable Market value.

3.5. Duplicate Data.

There are two instances where the same data is repeated in more than one place within the ‘CC STC Working Base Model’ spreadsheet. ‘BA Grads’ (appears in both the school type group worksheets and the School Data worksheets) and ‘Base Mission by School Type’ (appears in both the Working Model and School Data worksheets)⁴.

⁴ This data is not technically duplicated, since the values do no match. It appears that the values in the School Data worksheets have been adjusted from those shown on the Working Model worksheet. However, the data field is duplicated.

Chapter 4: Recommendations.

4.1. Overview of recommended changes.

In general terms, we recommend replacing the current implementation of two separate Excel spreadsheet files⁵ with a single Excel spreadsheet file containing four worksheets, and to restrict each workbook to a single academic year (using a filename that reflects the AY data contained in that file). The primary purpose of this change is to eliminate duplicate data within the files, reduce or eliminate errors due to manual data entry, and to eliminate the transition between the School Data worksheets and the ‘Working DST’ file altogether. The 14 worksheets in the “CC STC Working Base Model” file should be removed because they do not perform any unique function in the file (the summarizations performed in them can be handled by employing appropriate functions on a separate summary worksheet – this will be discussed in section 4.1.4 and 4.2.4). Additionally, this restructuring will result in a more modular arrangement by keeping data and reference tables separated from data processing and data summaries. The four worksheets can be thought of as modules that perform the following functions:

4.1.1. School Data Module (SDM).

This module is used for data storage only and will contain all the data for a school on (at a minimum) annual baccalaureate-level graduates, total enrollment, full-time enrollment, the number of non-US citizens, number of students over the age of 27, partnership data, and information on sister service competition. Most data in this module is either text or numeric, but the use of a couple of indicator (or Boolean – two-state – variables) is recommended. This also simplifies the data entry process, since all data pertaining to a particular school is contained in a single worksheet.

4.1.2. OCMPS Data Module (ODM).

This module contains a table with the figures for “Percent SALs” and “Percent First Stringers” assigned to their respective school type (from the OCMPS results). It is additionally recommended that the boundary values for academic reputation and school size are stored here. This will facilitate the ability to automatically look up a school’s values for “Percent SALs”, “Percent First Stringers”, and “Required Market Penetration” based solely on data stored in the

⁵ The file ‘CC STC Working Base Model’ has 19 worksheets (the ‘Main’ and “Definitions” worksheets were not included), and the ‘Working DST’ file has one worksheet.

School Data Module. This module would likely need modifications to its contents as the OCMPS progresses, and could be eliminated altogether once a percentage for SALs and First Stringers is determined for each school.

4.1.3. Mission Data Processing Module (MDPM).

This module is responsible for retrieving school data and performing the necessary calculations to determine a value for a commissioning mission for each school. It is where all the core functions of the model are executed. All columns in this module contain functions that either perform calculations on data stored in other modules (or workbooks), look up data stored in other modules (or workbooks), or some combination of these functions. The only column intended to store raw numerical data is a provision that allows and tracks deviations from the model's output. This aspect will be discussed in section 4.4.1.

4.1.4. Data Summary Module (SUM).

Any summarized data, such as the total number of schools within a region or brigade, other categorical totals, or grand totals, be stored here – separate from the raw data or other modules. There are a number of useful Excel functions that can be used to summarize data from anyplace in the workbook (or even data external to it, if needed).

4.1.5. Historical Data Module (HIST).

This can essentially be the same as the NINESTEP format currently in use. It is not strictly necessary to include this data as a module within the same workbook as the SDM, ODM, MDPM, and SUM modules. In fact, the preferred method would be to keep historical data in a separate workbook and simply refer to the appropriate cells as needed. The primary purpose for this module with respect to the calculation of the commissioning mission is to facilitate the implementation of the Alternate Minimum Mission Threshold that will be discussed in Section 4.4.2.

4.2. A Detailed Look at the Changes.

Figure 6 shows a diagram that represents our recommended implementation. The SDM is on top, and contains all the numeric data and indicator variables pertaining to each school. No functions are necessary or recommended on this worksheet. Note that the arrows pointing from the SDM to items in the MDPM represent functions in the MDPM that contain references back to the relevant cell in the SDM. The ODM contains the data shown in the diagram, but can also

contain the boundary values for school size and academic reputation that define a particular school type. The correct entry in the table can be accessed with the Excel's lookup functions. Finally, the MDPM performs all the actual calculations needed to deliver a value for commissioning mission.

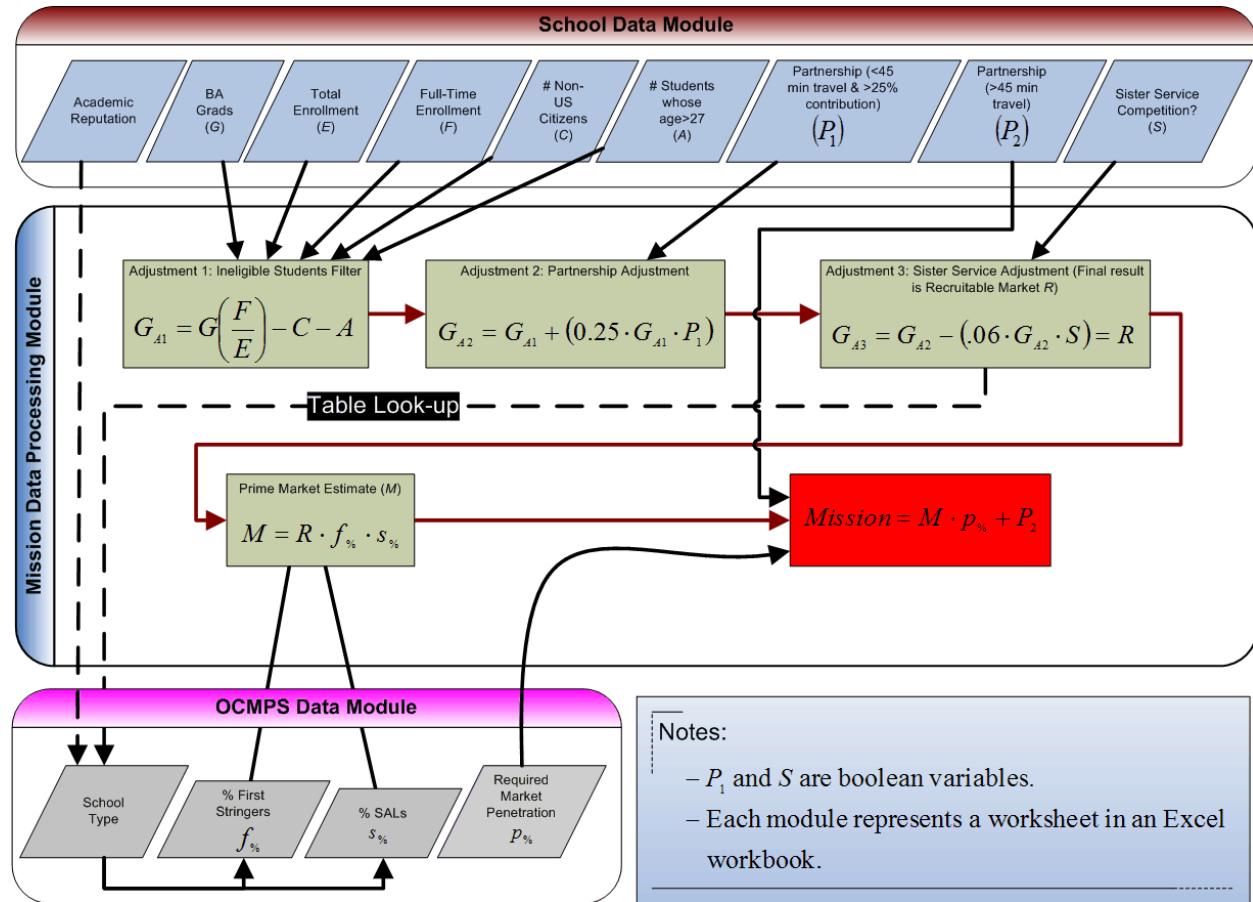


Figure 6: Proposed Model Implementation

4.2.1. Structure of the SDM.

As we stated before, the SDM should contain (at a minimum) columns for the data shown in Figure 6. It, of course, also may contain any other data that relates to a particular school (i.e. FICE, BDE, city, state, ZIP code, etc.). The primary function of this module is raw data storage. We recommend that any information that is derived from the raw data (i.e. school type from the OCMPS, such as “Small-Standard”) be reported in the MDPM using the appropriate function. The order of the data, however, is not important – any arrangement of the columns that makes data entry and review easier will work.

4.2.2. Structure of the ODM.

Our recommended structure for this module actually includes three tables: one table that reflects the academic reputation values, one that reflects the school size values, and one that assigns the SAL and First Stringer percentages to their respective school types. It can be laid out as shown in Figure 7.

	A	B	C	D	E	F
1			School Size Table			
2	Lower Bound	Size	Description			
3	0	Small	(defined as less than 800 undergraduate degrees)			
4	800	Medium	(defined as between 800 and 1499 undergraduate degrees)			
5	1500	Large	(defined as between 1500 and 2999 undergraduate degrees)			
6	3000	Mega	(defined as 3000 undergraduate degrees or more)			
7						
8		Academic Reputation Table				
9	Lower Bound	Reputation				
10	0	Standard				
11	2.5	Moderate				
12	3.2	High				
13	4	Prestigious				
14						
15		OCMPS Cluster Table				
16	Cluster	% 1st Stringers	% SALs	Required Market Penetration		
17	Small-Standard	13.4%	40.0%	18.1%		
18	Small-Prestigious	13.9%	51.0%	19.2%		
19	Small-Moderate	13.5%	40.0%	18.1%		
20	Small-High	14.6%	50.0%	12.8%		
21	Mega-Prestigious	15.3%	52.0%	12.9%		
22	Mega-High	17.5%	53.0%	12.9%		
23	Medium-Standard	13.5%	45.0%	12.8%		
24	Medium-Prestigious	14.9%	53.0%	12.9%		
25	Medium-Moderate	14.6%	44.0%	12.9%		
26	Medium-High	15.6%	47.0%	12.9%		
27	Large-Standard	13.2%	40.0%	12.9%		
28	Large-Prestigious	13.8%	49.0%	12.9%		
29	Large-Moderate	13.5%	42.0%	12.9%		
30	Large-High	14.8%	45.0%	12.9%		
31	MJC	50.0%	40.0%	15.6%		
32	SMC	50.0%	40.0%	15.1%		

Figure 7: OCMPS Data Module

The School Size table and the Academic Reputation Table are designed to be referenced using Excel's LOOKUP function, and the OCMPS Cluster Table is designed to be referenced using the VLOOKUP function. The advantage of this approach is that it will determine the appropriate school type, along with the figures for 'Percent First Stringers', 'Percent SALs', and 'Required Market Penetration' automatically, and any changes in the values stored in the ODM will be automatically reflected in the MDPM.

4.2.2.1. Getting School Type.

The School Size Table and Academic Reputation Table are referenced first in order to produce the appropriate school type label. This can be done in one step using a compound

function which utilizes the IF, LOOKUP, and CONCATENATE functions.⁶ In the example shown in Figure 8, we see that the school size is determined from taking the value in cell G2 (for the first school in the list) and using the LOOKUP function to retrieve the right category from the School Size Table in the ODM. Likewise, the school reputation is determined by taking the value for the school stored in the SDM and using LOOKUP to retrieve the correct category from the Academic Reputation Table. The result of the compound function then becomes the entry argument when referencing the OCMPS Cluster table.

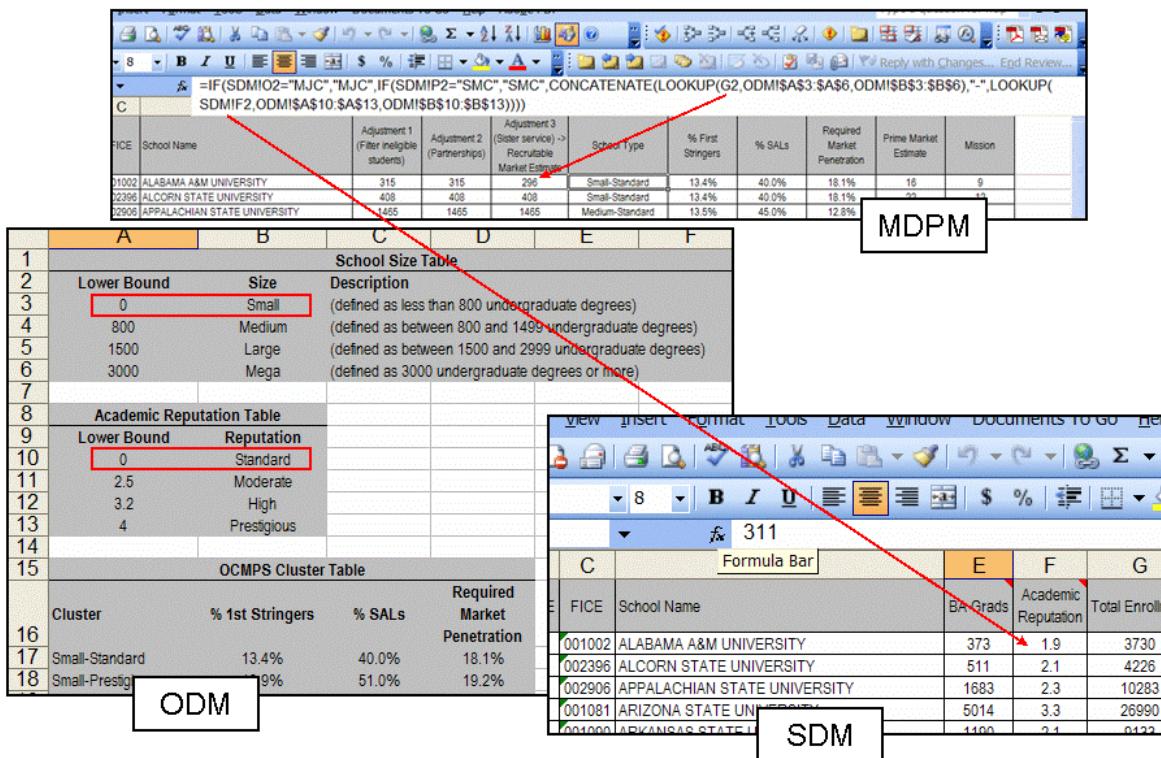


Figure 8: Getting School Type from the tables in the ODM.

4.2.2.2. Getting OCMPS Percentages.

Once a school type has been determined, the OCMPS percentages (Percent First Stringers and Percent SALs) for that type can be found using the VLOOKUP function. An example that finds “% First Stringers” is shown in Figure 9.

⁶ The function first checks to see if the school is a MJC or SMC. If not, then the two LOOKUP commands locate the appropriate size and reputation, respectively, and the CONCATENATE function places them in a single string, joined with a hyphen.

The screenshot shows a Microsoft Excel spreadsheet with two tables. The top table is a list of schools with various metrics. The bottom table is a 'OCMPS Cluster Table' with columns for Cluster, % 1st Stringers, % SALs, and Required Market Penetration. A red box highlights the formula '=VLOOKUP(H2, ODM!\$A\$17:\$D\$32, 2, FALSE)' in the top table's formula bar, and another red box highlights the 'Small-Standard' entry in the cluster table.

C	D	E	F	G	H	I	J	K
FICE	School Name	Adjustment 1 (Filter ineligible students)	Adjustment 2 (Partnerships)	Adjustment 3 (Sister service) -> Recruitable Market Estimate	School Type	% First Stringers	% SALs	Required Market Penetration
01002	ALABAMA A&M UNIVERSITY	315	315	296	Small-Standard	13.4%	40.0%	18.1%
02396	ALCORN STATE UNIVERSITY	408	408	408	Small-Standard	13.4%	40.0%	18.1%
02906	APPALACHIAN STATE UNIVERSITY	1465	1465	1465	Medium-Standard	13.5%	45.0%	12.8%
01081	ARIZONA STATE UNIVERSITY	4419	5524	5524	Mega-High	17.5%	53.0%	12.9%
01090	ARKANSAS STATE UNIVERSITY	996	996	996	Medium-Standard	13.5%	45.0%	12.8%
08310	AU BEMERY	508	508	508	Small-Standard	13.4%	40.0%	18.1%

12	3.2	High		
13	4	Prestigious		
14				
15	OCMPS Cluster Table			
16	Cluster	% 1st Stringers	% SALs	Required Market Penetration
17	Small-Standard	13.4%	40.0%	18.1%
18	Small-Prestigious	13.9%	51.0%	19.2%
19	Small-Moderate	13.5%	40.0%	18.1%
20	Small-High	14.6%	50.0%	12.8%
21	Mega-Prestigious	15.3%	52.0%	12.9%
22	Mega-High	17.5%	53.0%	12.9%
23	Medium-Standard	13.5%	45.0%	12.8%
24	Medium-Prestigious	14.9%	53.0%	12.9%
25	Medium-Moderate	14.6%	44.0%	12.9%
26	Medium-High	17.0%	47.0%	12.9%
27	Large Standard	40.0%	40.0%	12.0%

Figure 9: Getting OCMPS Percentages from the ODM

Here, the type “Small-Standard” becomes the first argument (the lookup value) in the VLOOKUP function. The function tries to match the lookup value with an entry appearing in the left-hand column of the table (which is column 1). If a match is found, it returns the value stored in a cell whose row is the same as the lookup value and whose column is identified in the third argument of the VLOOKUP function (2, in this case). The values for “% SALs” and “Required Market Penetration” can be found similarly – the only change is the third argument in the VLOOKUP function (which would change to 3 and 4, respectively). The “Required Market Penetration” percentages shown in Figure 9 were taken directly from the wargaming section of the “Working Model” worksheet in “CC STC Working Base Model” spreadsheet. They are applied in the Mission calculation without modification. This ensures a consistent, systematic approach to determining the commissioning mission values. Deviations from these values (via negotiation) are accounted for using one of the methods outlined in Section 4.4.1.

4.2.3. Structure of the MDPM.

The primary function of this module is to produce the commissioning mission values. There should be no need to enter or store any raw data here – all cells on this worksheet should either contain direct references to cells on other worksheets (when data is “being brought forward” with no calculations performed, such as a school’s name), or perform a needed calculation or other process on the data. An example of a basic layout that reflects the structure shown in Figure 6 is now shown in Figure 10. Cells in the “School Type” column use the compound function shown in Figure 8 and discussed in section 4.2.2.1. Cells in the “% First Stringers”, “% SALs”, and “Required Market Penetration” columns use the function shown in Figure 9 and discussed in section 4.2.2.2. The order of the columns in this example reflects the process shown in Figure 6, where we adjust BA Grads to get the Recruitable Market and use that value to find the right School Type. We use School Type to find the right OCMPS percentages, which are then used in the calculations for Prime Market Estimate and Mission.

	A	B	C	D	E	F	G	H	I	J	K	L	M
Region	BDE	FICE	School Name	Adjustment 1 (Filter ineligible students)	Adjustment 2 (Partnerships)	Adjustment 3 (Sister service) -> Recruitable Market Estimate	School Type	% First Stringers	% SALs	Required Market Penetration	Prime Market Estimate	Mission	
1													
2	W	8	001002	ALABAMA A&M UNIVERSITY	315	315	296	Small-Standard	13.4%	40.0%	18.1%	16	9
3	W	8	002396	ALCORN STATE UNIVERSITY	408	408	408	Small-Standard	13.4%	40.0%	18.1%	22	12
4	E	5	002906	APPALACHIAN STATE UNIVERSITY	1465	1465	1465	Medium-Standard	13.5%	45.0%	12.8%	89	12
5	W	14	001081	ARIZONA STATE UNIVERSITY	4419	5524	5524	Mega-High	17.5%	53.0%	12.9%	512	66
6	W	10	001090	ARKANSAS STATE UNIVERSITY	996	996	996	Medium-Standard	13.5%	45.0%	12.8%	61	13
7	W	8	006310	AUBURN UNIV AT MONTGOMERY	508	508	508	Small-Standard	13.4%	40.0%	18.1%	27	8
8	W	8	001009	AUBURN UNIVERSITY	2034	2543	2390	Large-Moderate	13.5%	42.0%	12.9%	136	18
9	F	5	001562	AIKEN STATE UNIVERSITY	527	650	650	Small-Standard	13.4%	40.0%	18.1%	35	6

Figure 10: An example MDPM. Note that the school name is linked over from the SDM.

4.2.3.1. Getting the Recruitable Market Estimate.

Functions stored in columns E, F and G perform the calculations necessary to generate an estimate for “Recruitable Market” (the equations for each step are shown in Figure 6). The progression on the worksheet is left-to-right. All of the data used for the “Adjustment 1” (Ineligible Student Filter) calculation is referenced from the SDM. This adjustment results in an interim value (G_{A1}) that is no greater than the BA Grads value. The ROUND function is employed to produce an integer value (see Figure 11). Notice also that all ineligible students (part-time, overage, non-citizens) are removed from the BA Grads value in a single calculation.

The figure shows two Microsoft Excel spreadsheets side-by-side. The top spreadsheet contains a table with columns labeled C through H. The formula in cell E1 is =ROUND(SDM!E2*(SDM!H2/SDM!G2)-SDM!I2-SDM!J2,0). The bottom spreadsheet contains a table with columns labeled D through J. The formula in cell E1 is =ROUND(311,0). Red arrows indicate data flow from the 'School Name' column of the top table to the 'School Name' column of the bottom table, and from the 'Market Estimate' column of the top table to the 'Non-citizens' column of the bottom table.

Figure 11: Calculating Adjustment 1.

The calculation for “Adjustment 2” (Partnership Adjustment) uses the previous result (G_{A_1}) and additional data from the SDM to produce another interim result, G_{A_2} . This value will always be no less than G_{A_1} . Note also that this function employs one of our Boolean variables (P_1 , which is 1 if a partnership school is within 45 minutes of the program school and contributes at least 25% of the programs enrolled cadets, and 0 otherwise). If the value is zero, $G_{A_2} = G_{A_1}$.

The figure shows two Microsoft Excel spreadsheets side-by-side. The top spreadsheet contains a table with columns labeled C through G. The formula in cell E1 is =ROUND(E2+(0.25*E2*SDM!Q2),0). The bottom spreadsheet contains a table with columns labeled D through R. The formula in cell E1 is =ROUND(311,0). Red arrows indicate data flow from the 'School Name' column of the top table to the 'School Name' column of the bottom table, and from the 'Market Estimate' column of the top table to the 'Near Partnership' column of the bottom table.

Figure 12: Calculating Adjustment 2.

The final adjustment to BA Grads is Adjustment 3 (Sister Service Adjustment) which results in the Recruitable Market Estimate, R . The adjustment employs our other Boolean variable, S , which is 1 if another ROTC program exists on campus, and 0 otherwise. It has the same effect on this calculation as P_1 did in Adjustment 2. The value for R will never be greater than G_{A2} .

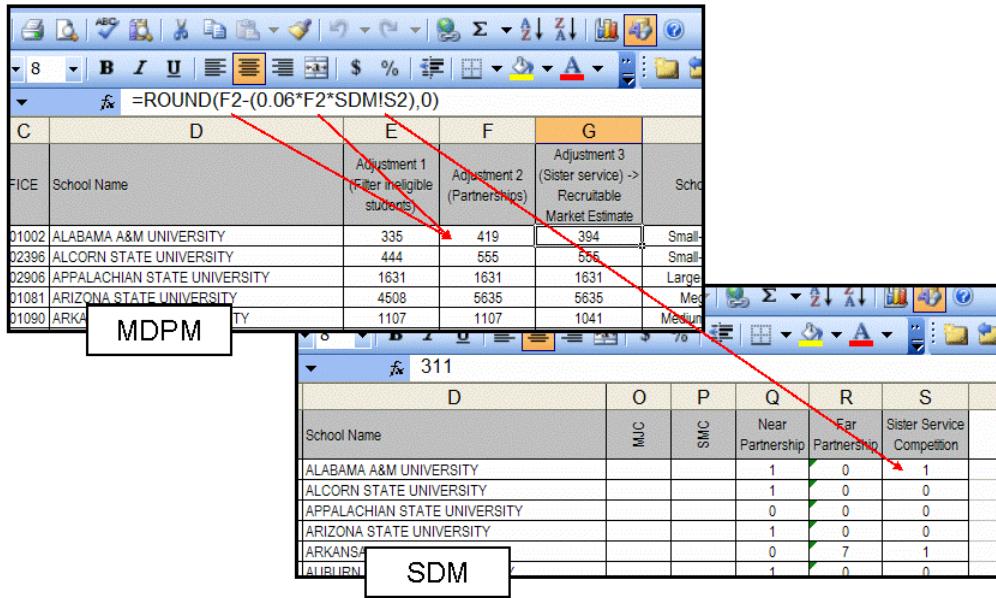


Figure 13: Calculating Adjustment 3. This result is the Recruitable Market Estimate

4.2.3.2. Getting the Prime Market Estimate.

The “Prime Market Estimate” calculation is just the product of the values in columns G, I, and J. All of these values are already present in the MDPM, so there are no references to other worksheets in the formula.

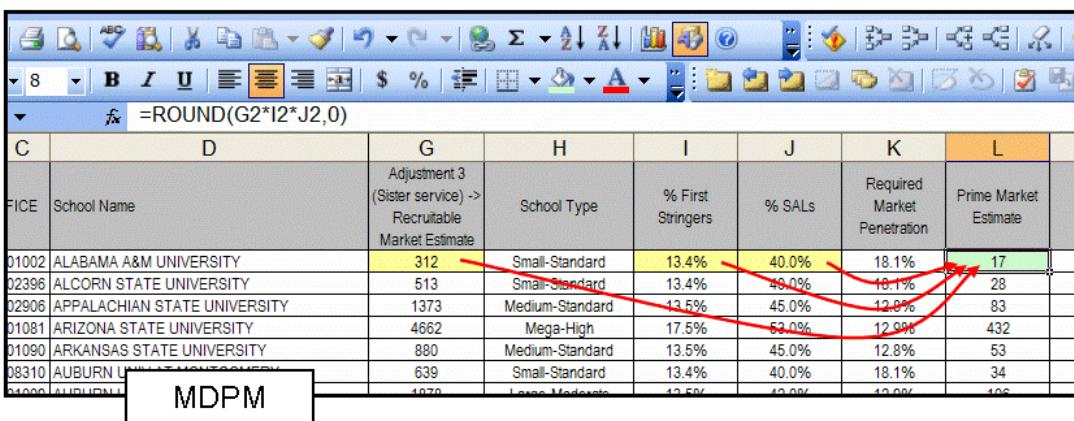


Figure 14: Calculating the Prime Market Estimate.

This is the only time in our implementation that an estimate of the Prime Market is derived.

4.2.3.3. Getting the Commissioning Mission.

The value for “Mission”, as shown in the formula in Figure 6, is the product of the Prime Market Estimate and the Required Market Penetration, plus any cadets resulting from a partnership school outside the 45-minute radius. An example of the final calculation is show in Figure 15.

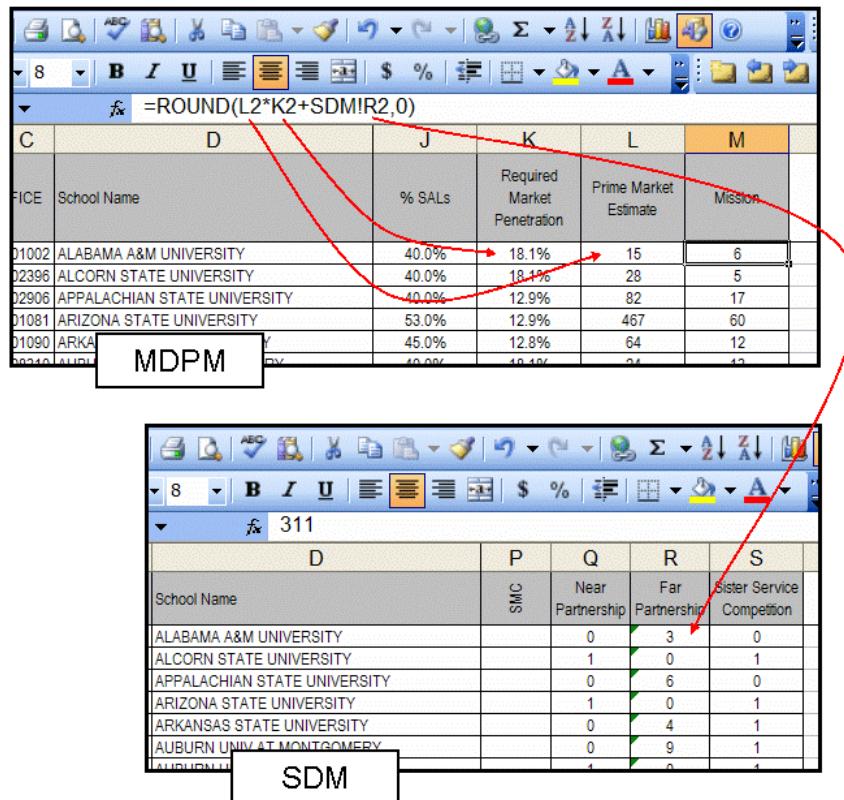


Figure 15: Calculating the Commissioning Mission.

4.2.4. Suggested Functions for SUM.

There are several useful functions that can be employed in the SUM module to provide summary information on the data found in other modules. Some of the most commonly used functions include COUNT, COUNTIF, and SUMIF (consult the Excel Help files for details on their use). By way of example, we will show a simple table that provides counts of the different school types in a compact format. In this case we will use both the COUNTIF and CONCATENATE functions. In Figure 16, we show how a compound function using these two built-in functions can be used to provide a count of the number of schools in each category.

The syntax of the COUNTIF function requires a range of cells in the first argument, and the counting criterion in the second argument. It is in the second argument that we use the CONCATENATE function.

	A	B	C	D	E	F	G	H	I
1		Small	Medium	Large	Mega				
2	Standard	33	20	18	0	71			
3	Moderate	21	24	46	1	92			
4	High	17	15	19	18	69			
5	Prestigious	6	11	5	7	29			
6		77	70	88	26				

Figure 16: An example of a school type summary.

The first argument in the COUNTIF function refers to the column on the MDMP that contains the school types (size-reputation). In cell B2, we have built the criterion by using the CONCATENATE function to create a string from the contents of cells B1 and A2, separated by a hyphen (the result of which is “Small-Standard”). It then returns the number of schools that have “Small-Standard” stored in the school type column. The dollar signs used in the CONCATENATE terms ensure that the function continues to reference the cells where the labels are stored after it is copied to the other 15 summary cells.⁷

4.3. A Word about Inputs.

We have seen no indications of any problems with the model’s current inputs. Since the OCPMS is ongoing, the percentages of First Stringers and SALs will eventually be derived from survey data at the school level and will no longer need to be aggregated across school types. As for propensity, when the First Stringer criterion is compared with measures of propensity in the DoD’s Youth Attitude Tracking Study (YATS), the approaches are similar (1995 Annual Defense Report: Appendix G). YATS usually asks respondents (who are still in high school) directly about their intention to enlist (unless they provide an “unaided mention” of intended military service (Orvis, 2001)), which is essentially a direct question about their plans for the future (something that senior highs school students tend to think about). This direct approach is unlikely to have the same effect on a first or second year college student, since at that point they “have already embarked” on their future plans. Because of this, we think that the OCMPS takes

⁷ Please note that the SUM label that appears in the CONCATENATE terms refers to the name of the worksheet that the summary table is stored on, and not Excel’s built-in SUM function.

a reasonable approach for determining propensity by determining whose attitudes and interests align with the activities of Army ROTC and Army service.

4.4. Additional Modifications.

There are a couple of other tools that could be added to this basic implementation.

Section 4.4.1 is a modification to an existing calculation in the current STC implementation. Primary emphasis in this section is given to the Alternate Minimum Mission Threshold in Section 4.4.2.

4.4.1. Delta to Track Mission Changes.

We expect model fidelity to increase as more data from the ongoing OCMPS is introduced into the model (as well as other future changes discussed in Chapter 4). However, a method is needed to allow changes to the commissioning mission values without disturbing the functions stored in existing cells (these changes are the result of the current policy discussed in section 1.4). The simplest method of doing this would be to add two columns to the right of the “Mission” column in the MDPM: the first column represents the “Delta” (the increase or decrease in the “Mission” value), and a second column called “Adjusted Mission”, which is just the sum of the values for “Mission” and “Delta”. An example is shown in Figure 17.

M	N	O
Mission	Delta	Adjusted Mission
12		=M2+N2
12		

Figure 17: Using a delta to adjust "Mission" values.

This also allows the changes in the “Mission” values to be explicitly tracked and stored without overwriting the previous equations. The “Adjusted Mission” could also be calculated by changing the “Required Market Penetration”. If this approach is preferred, it is recommended that the “Delta” column (which could be renamed “New Required Market Penetration”) contain a function that produces the required percentage value, and “Adjusted Mission” would contain constants. If we let $np_{\%}$ equal “Required New Market Penetration”, then

$$\frac{\text{Adjusted Market} - P_2}{M} = np\%$$

Equation 2: Equation for expressing delta as a new required market penetration percentage.

4.4.2. Alternate Minimum Mission Threshold Derived from Production History.

In the current STC implementation, if the model produces a commissioning mission value less than six, the value is increased to six. USACC has decided that this should be the absolute minimum commissioning mission assigned to any school. Since the overall commissioning mission is increasing (3900 to 4400), we examined historical production data that we were provided for insights on any other potential minimum mission assignments. The goal of this portion of the analysis was to increase the chances of achieving the new overall commissioning mission of 4400 officers.

4.4.2.1. Initial Observations.

Production data for 1991-2004 was examined. The 1990 data was excluded because the overall commissioning mission was significantly higher that year than in subsequent years. The 1991-2004 data for each school was tested for normality using a Ryan-Joiner test. This test is similar to using a normal probability plot to determine whether observed data could have been taken from a normal population (an approximation in this case because our data is actually discrete, whereas the normal distribution is continuous). However, the Ryan-Joiner test is a quantitative measure, which allowed us to test all the schools at once.⁸ Under this test, the null hypothesis was rejected for seven schools at a significance level of $\alpha = 0.01$ (which corresponds to a sample correlation coefficient of $r = 0.9049$). Of those seven schools, none had a sample correlation coefficient lower than 0.86. Figure 18 shows the distribution of sample correlation coefficients for 270 of the schools (two schools were removed due to insufficient data).

⁸ For an explanation of the Ryan-Joiner test, see Appendix A.

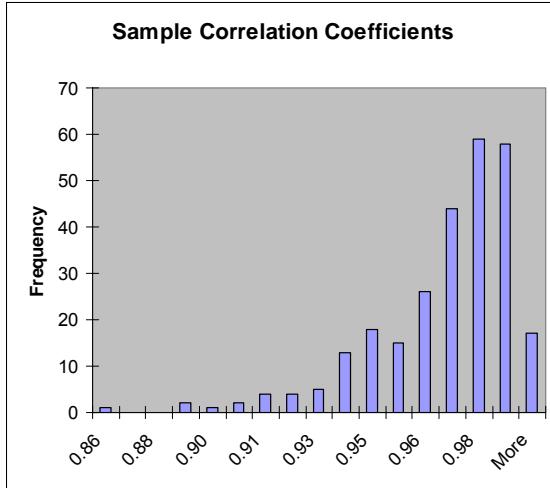


Figure 18: Histogram of sample correlation coefficients from the Ryan-Joiner test.

Since the sample size for each school is relatively small, we concluded that the normal distribution is a good approximation for the data. We used the sample mean and standard deviation for each school as estimates for the normal distribution's parameters μ and σ .

4.4.2.2. Methodology.

A simulation was conducted that allowed each school to generate random commissioning mission values from the normal distributions that were estimated from production data. However, a limitation was imposed on each distribution as to *where* those random numbers could be drawn from. The point here was to determine if there was a minimum percentile, applied uniformly to each school's distribution, which would result in a total production of 4400 most of the time. To illustrate this, we'll examine one school, Alabama A&M University. This school had a production sample mean of 12.36 and a sample standard deviation of 4.91. If we allow the simulation to generate random values from the entire normal distribution, the vast majority of the values will be in the interval (-2.37, 27.09), and about 2/3 of the values will be in the interval (7.45, 17.27). Figure 19 shows the distribution's pdf, and the histogram on the right shows simulation results from 1000 samples.

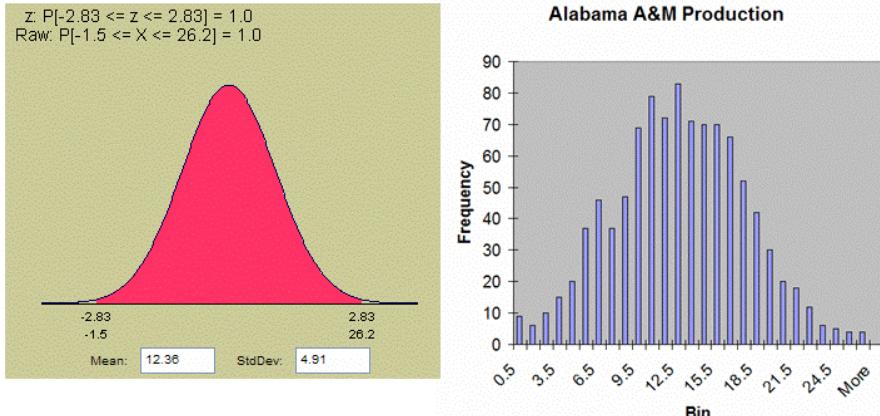


Figure 19: Simulating school production from the whole distribution.

According to USACC policy, no school will be assigned a commissioning mission lower than six. If we set this as the lower limit for Alabama A&M, we get the following results:

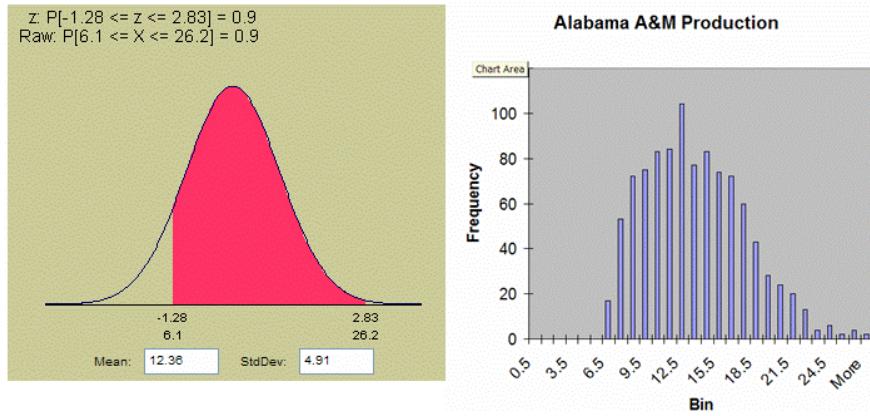


Figure 20: A lower threshold of 6 is imposed.

This doesn't appear to set the bar very high for Alabama A&M, since the graphs in Figure 20 imply that they would have no problem achieving six commissions.

At this point, it is appropriate to step back and examine the overall goal once more. In the current model, a school is assigned either a commissioning mission of six, or the mission value produced by the model, whichever is greater. This rule was applied as the only constraint in our simulation, so that a school was allowed to generate random production values from its distribution, but never lower than six. Using this rule, the simulation almost never generated a total production in excess of 4400 out of 500 trials. Based on this result, our goal was to find an additional constraint: Assign a school the greatest of three possibilities: (1) the absolute minimum of six, (2) the model output, or (3) the n^{th} percentile of a school's distribution. The goal was to find a value for n (to be applied to *all* schools) such that the total production was at

least 4400 for 95% or more of the simulation trials. In our simulation, selecting the 34th percentile as the additional constraint produced the desired result. In other words, if all schools in the simulation generated random production numbers consistent with their distributions, but bounded below by either the constant 6 or the distribution's 34th percentile (whichever was greater), the simulation produced a total of 4400 commissions or greater at least 95% of the time. When applied to Alabama A&M, we get the results shown in Figure 21.

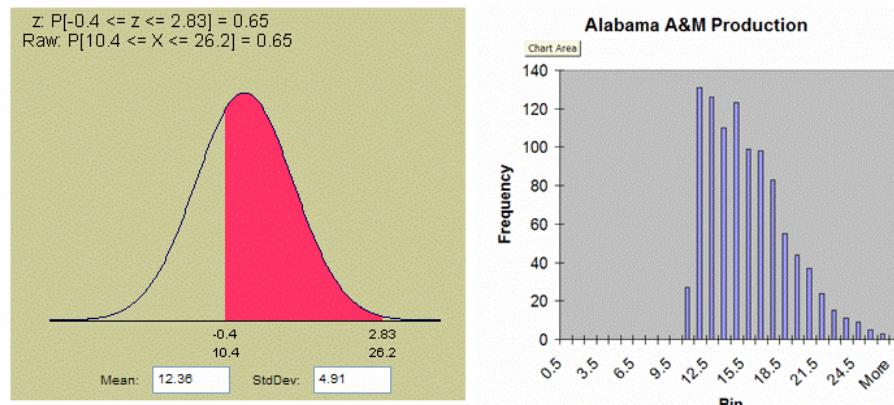


Figure 21: The lower threshold is set at the 34th percentile.

In this case, the 34th percentile is greater than 6, so it would be used as the minimum for this school. If the model output for Alabama A&M turns out to be greater than 10 (rounding the 34th percentile to the nearest value), it would be the commissioning mission assigned. Otherwise, 10 would be the assigned mission. For more detailed information on the simulation, see Appendix B.

4.4.2.3. An Important Note on Assumptions.

In reality, a school will rarely produce the exact number of officers in a given year as the commission mission. It was assumed that increasing the mission of a school (at least by relatively small amounts) would, on average, result in a production increase. This is important to our simulation since the distributions are based on production history, but we are using the results to justify a change in mission values. So, to check this, we performed a simple regression on a subset of the data from 1991 to 2004.⁹ The results showed that, on average, an increase in mission by one corresponded to an increase in production of about 0.95 (this value was the

⁹ A total of 224 schools were examined. Schools were excluded from this analysis because they either had missing data or were known outliers, such as the MJC's and SMC's.

average slope from 14 separate regression analyses; the standard deviation of this average slope was about 0.12). This implies that if a school's mission increased by a small amount, on average it will meet the increase in mission.

4.4.2.4. Calculation of the Percentile.

For this new criterion to be included in the mission assignment process, the 34th percentile needs to be calculated for each school. There are two ways this can be done, but the results may be slightly different (the magnitude of the difference is related to how closely the data was approximated by the normal distribution; in our examination of the data, 220 of the schools showed a difference of one or less when the two methods were compared).

4.4.2.4.1. Directly from Historical Data.

A percentile can be calculated directly from a dataset using the PERCENTILE function in Excel. The syntax of this function is PERCENTILE(array,k), where 'array' is a cell range that contains the data, and 'k' is the percentile expressed as a probability (between 0 and 1).

4.4.2.4.2. From the Approximate Normal Distribution.

This can be done using the NORMINV function. The syntax for this function is NORMINV(probability,mean,standard_dev), where 'probability' is a cumulative probability value, 'mean' is the arithmetic mean of the distribution, and 'standard_dev' is the standard deviation of the distribution. In this case, our cumulative probability value is 0.34, and we substitute the sample mean and sample standard deviation from each school's production data as the estimates for the distribution parameters. The function returns the 34th percentile of the distribution.

4.4.2.5. Implementation.

Figure 22 shows examples of both of the methods explained in the previous sections. In these examples, the desired percentile (34th) is entered as a constant for the appropriate argument. If the user wants the flexibility to change this value, it should be stored elsewhere and referred to in the function used to calculate the percentile.

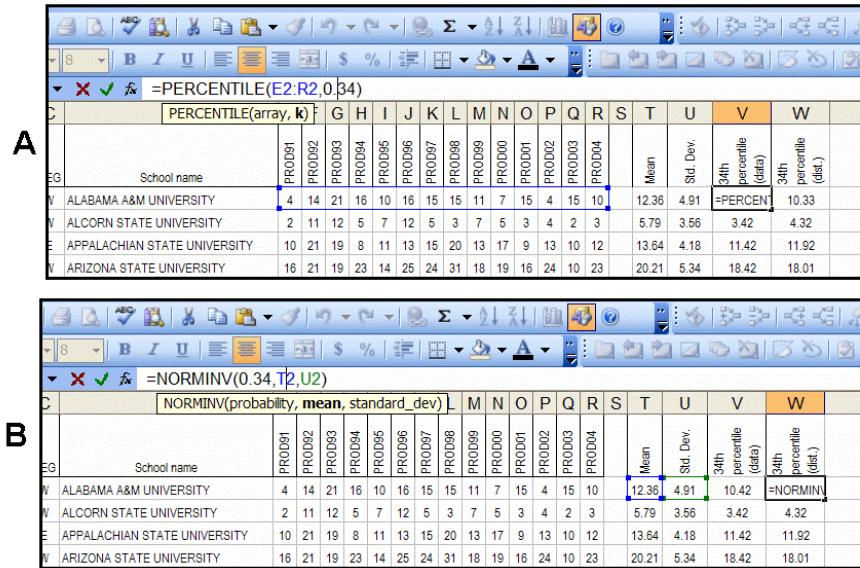


Figure 22: Examples of percentile calculation.

Example A in Figure 22 shows the use of the PERCENTILE function to calculate the 34th percentile from the 1991-2004 production data. Example B shows the use of the NORMINV function to calculate the 34th percentile from the sample mean and sample standard deviation. Once the 34th percentile is determined, it can be incorporated in the decision for commissioning mission value by using the MAX function. The most concise way to do this would be to replace the function in the ‘Mission’ column of the MDPM (see Figure 15) with the MAX function, with the existing function entered as one of the arguments. An example of how to do this is shown in Figure 23.

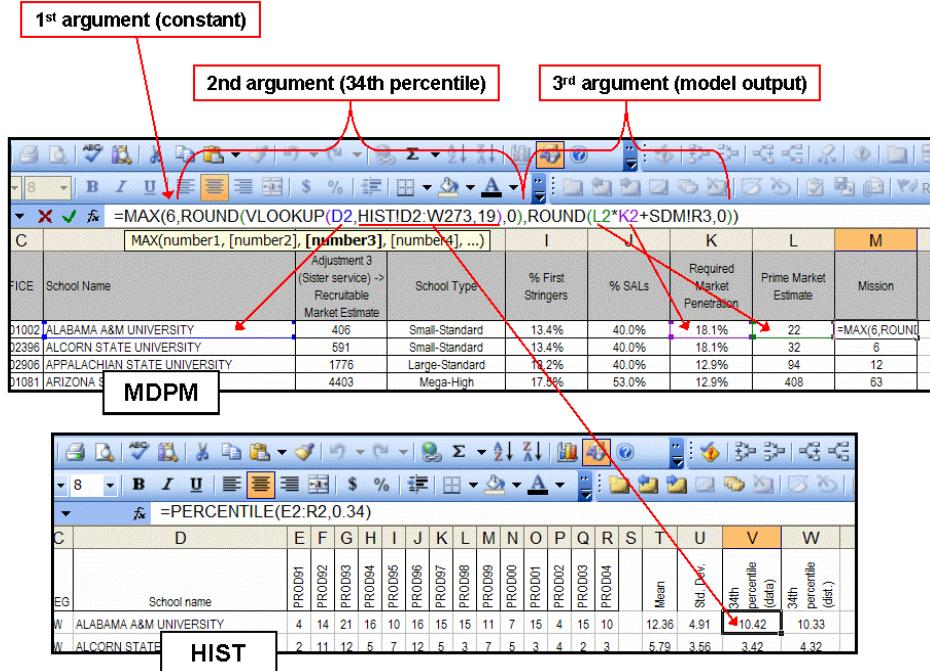


Figure 23: Using the Alternate Minimum Mission Threshold.

Here, the first argument in the MAX function is the absolute minimum of 6, entered as a constant. Our 34th percentile value is the second argument. Note that we have used the VLOOKUP function here to find the percentile value that we calculated in the HIST module. This was done so that, if the data in the HIST module is sorted differently than the order of schools in the MDPM, we will still be assured of getting the right percentile.¹⁰ The third argument is the model output – the same function that was used before (see Figure 15). In this example, we have used the ROUND function to round the results of the functions used in the second and third arguments to the nearest integer. This compound function produces the desired result – it sets the mission value at either 6, the 34th percentile of the particular school, or the model output of the particular school, whichever is the greatest.

4.4.3. Comparing Mission Numbers to the Estimated Production Distributions.

Since normal distributions were used to estimate the production distribution for each school, those distributions can be used to provide a measure of an assigned commission mission value (either from the model or from negotiation) with respect to the school's production history.

¹⁰ The table that VLOOKUP refers to in this example has 'School Name' as its first column (column 1), making the column containing the 34th percentile (from data) column 19.

Also, since the parameters for each school's distribution have been estimated from the data, the **NORMDIST** function can be used to determine where a mission value 'fits' within a school's production history. The syntax for this function is **NORMDIST(x, mean, standard_dev, cumulative)**, where **x** is the mission value we want to evaluate, **mean** is the sample mean from the production data, **standard_dev** is the standard deviation from the production data, and **cumulative** is a Boolean variable (when set to **TRUE**, the function returns the cumulative probability $P(X \leq x)$). This function can quickly be applied to each school, and the results used to determine if any mission values (either from model output or post-negotiation) appear extremely unlikely when compared to a school's production history (indicated by cumulative probabilities either near zero or one). These schools can be checked again to ensure that they are receiving realistic mission values.¹¹

¹¹ It should be noted that this is intended for evaluation only. For example, a school's mission value may be very high compared to its production history, resulting in a cumulative probability near one, but there may be a compelling reason for the high mission value, such as a much larger recruitable market than was previously known, for which that school's ROTC program will be receiving additional resources to facilitate greater market penetration. Hence, the primary reason for this tool is simply to identify schools that receive mission values considered extreme when compared to their production history.

Chapter 5: Performance Measures.

5.1. Measuring Efficiency with Data Envelopment Analysis (DEA).

The complexity of the commissioning process and the number of factors that contribute to the successful commissioning of a certain cadet makes the use of simple ratios to represent efficiency of very limited value. There are several factors that could potentially affect the number of commissioned officers produced by a certain school. These include the assigned commissioning mission, the number of cadre, operating budgets, scholarships, market characteristics, space constraints, advertising, and so on. In simpler problems, a ratio of output to input can be used as a measure of efficiency. In our problem, we need to be able to account for all the relevant inputs and outputs when attempting to measure efficiency. Data Envelopment Analysis (DEA) makes this possible. Unlike regression analysis, which compares each unit of interest against the central tendency of the whole, DEA finds the best performers based on the inputs and outputs provided and compares each one against the best performers (strictly speaking, against a reference set that is identified for each decision-making unit, or DMU). Furthermore, DEA does not require the assignment of fixed weights to the various inputs (the values of which are typically hard to justify) – weights that are then identically applied to each DMU. Instead, the weights are assigned from the data, and each DMU is assigned a “best set” of weights. This method is a much more comprehensive approach to separating efficient and inefficient performers, and is generally perceived as a fairer approach by the units being examined.

5.2. The Purpose of This Chapter.

It is not our intent here to proceed with a detailed explanation of DEA. For a more detailed explanation of DEA, see (Cooper, 2000). What we will primarily do here is work through an example dataset using a DEA software application called Frontier Analyst Professional, in order to demonstrate how it can be used to provide useful information regarding the efficiency of ROTC programs in the near term. An example of a model that incorporates DEA with production functions can be found in (Brence, 2004).

5.3. An Example Using Frontier Analyst Professional.

To demonstrate the process used in DEA, we will perform an analysis on an example dataset using Frontier Analyst Professional from Banxia Software. The dataset used in this

example was derived from three columns of data stored in the “CC STC Working Base Model” spreadsheet: Under “Historic Production Data”, the data in the “3-Yr Avg Cmsn Msn” and “3-Yr Avg Prod” columns, and under “Historic Cadre”, the data in the “Avg # of Cadre” column. This analysis consists of three basic steps (adapted from the Frontier help documentation):

- 1) Defining and selecting the units (DMUs) to use in the analysis.
- 2) Deciding which factors to use for inputs and outputs.
- 3) Running the analysis and interpreting the results.

In our example, the schools represent our DMUs, and we will filter the dataset to exclude all HBCUs, SMCs, and MJCs. The inputs we will use for the initial analysis are “3-Yr Avg Cmsn Msn” and “Avg # of Cadre”, and the single output will be “3-Yr Avg Prod”.

5.3.1. Getting Data into Frontier Analyst.

There are several ways to get your data into Frontier Analyst (as can be seen in the dialog shown in Figure 24), but the method we will use is to simply paste it in from the clipboard (which is done by selecting the first option as shown):

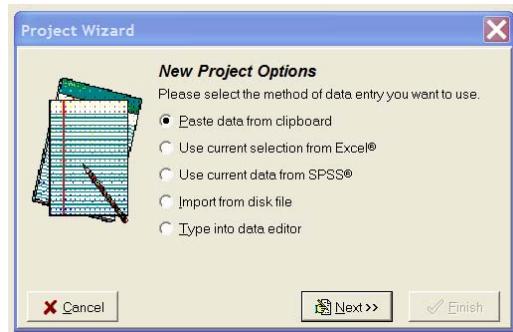
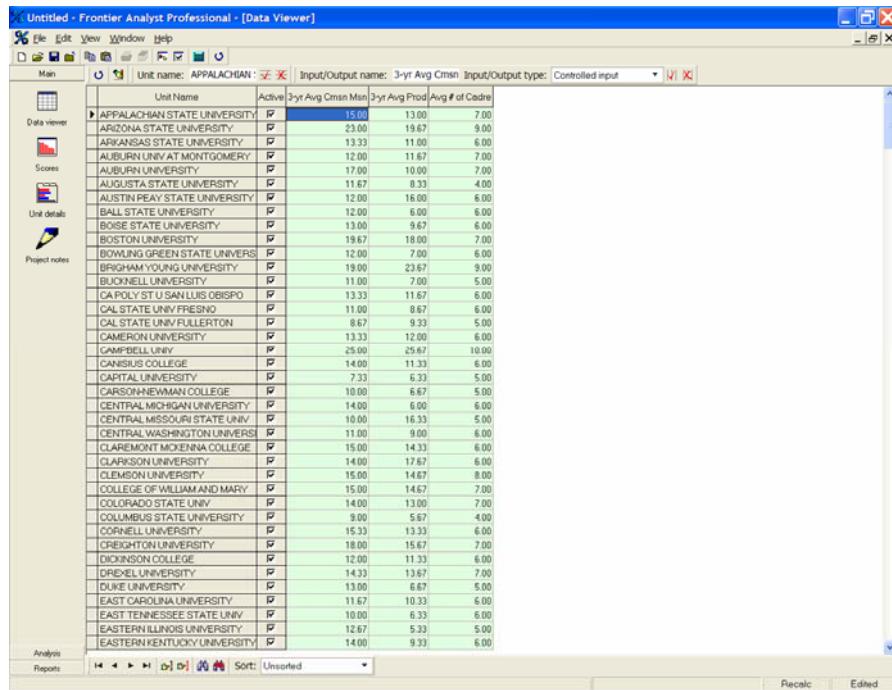


Figure 24: Data Input Dialog

For this example, we filtered the data in the spreadsheet (using AutoFilter) to remove the HBCUs, MJC, and SMCs, then copied the data to the clipboard.¹² This removed from our example dataset the primary known outliers (there will be a small group of other outliers that we will remove individually from the dataset in a later step). Once that is done, we clicked ‘Next’ here and in all subsequent dialogs (since no other changes were needed) until we reached the end of the dialog. Once finished, Frontier Analyst opens the Data Viewer, which contains our

¹² Actually, there is an additional step here. Since the data columns in Excel are not contiguous, they were copied and pasted into a blank worksheet, then copied from there and pasted into Frontier. Otherwise we got an error stating we were pasting too many variables.

dataset. From here we will be able to examine the dataset and make decisions as to which variables need to be identified as inputs (controlled or uncontrolled) and outputs.



The screenshot shows the 'Data Viewer' window of Frontier Analyst Professional. The window title is 'Untitled - Frontier Analyst Professional - [Data Viewer]'. The menu bar includes File, Edit, View, Window, Help. The toolbar has icons for New, Open, Save, Print, and others. The main area displays a table with the following columns: Unit Name, Active, 3-yr Avg Cmsn Min, 3-yr Avg Prod, and Avg # of Cedre. The table lists numerous universities, each with a checked 'Active' status. The '3-yr Avg Prod' column contains values such as 15.00, 23.00, 13.33, etc. The 'Avg # of Cedre' column contains values such as 7.00, 9.00, 6.00, etc. The bottom of the window shows a toolbar with buttons for Analysis and Reports, and a status bar indicating 'Sort: Unsorted'.

Figure 25: STC data in the Data Viewer.

The data that we pasted into Frontier Analyst included the school names along with the three variables of interest. The data can also be edited in this window directly, if needed.

5.3.2. Setting Inputs and Outputs.

By default, Frontier Analyst identifies all the variables entered into the data viewer as controlled inputs¹³. Variables can also be identified as uncontrolled inputs or as outputs. An uncontrolled input would be an input that we have no power to increase or decrease (an example from business might be the number of competitors in a given area). For this example we assume that all input variables are controlled inputs, so we will not change them. However, we need to change the variable “3-Yr Avg Prod” to an output. To do this, you simply click in any one of the cells in that column and then change the Input-Output type for that variable by selecting “output” from the pull-down menu (see Figure 26).

¹³ Controlled and uncontrolled inputs are referred to discretionary and non-discretionary inputs in (Cooper, 2000:63).

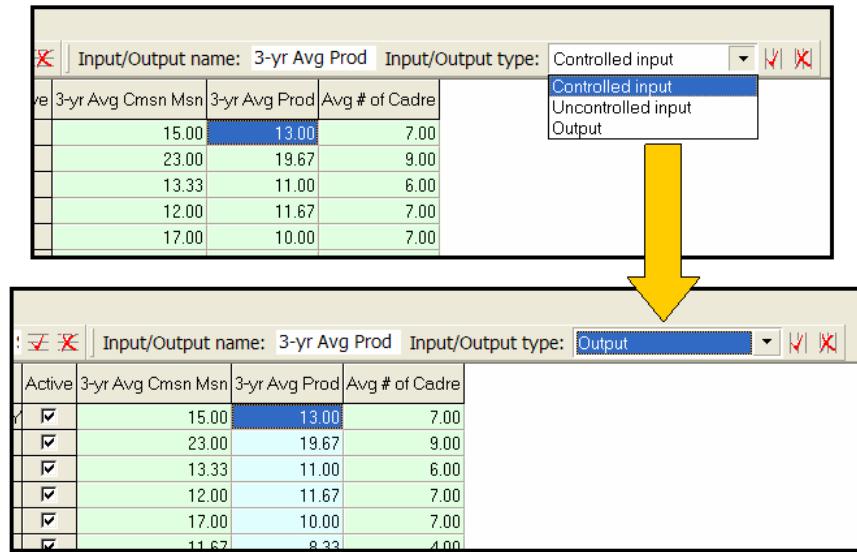


Figure 26: Identifying the output variable.

The column background color changes to identify it as an output variable. There were also three specific schools that we decided to remove from the dataset (they were removed because they contained limited historical data in the 1991-2004 data). This was done directly in the Data Viewer by selecting the schools and then clicking the “Deactivate Current Unit” button:

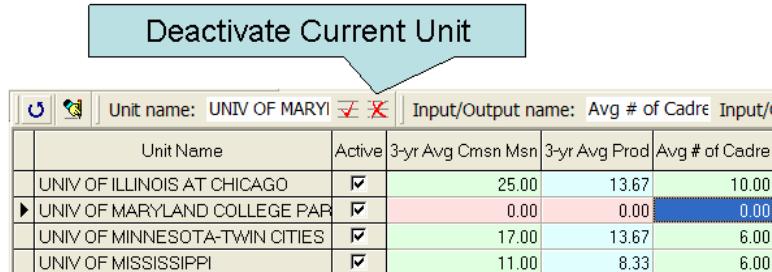


Figure 27: Removing an individual school from the dataset.

This action causes these DMUs to be removed from the analysis (their removal is indicated by graying out the text – they are not deleted from the data viewer). Once all modifications to the data have been done, we need to decide what type of DEA analysis we want to perform with respect to optimization type and scaling type. These selections are made through the DEA options dialog (shown in Figure 28).

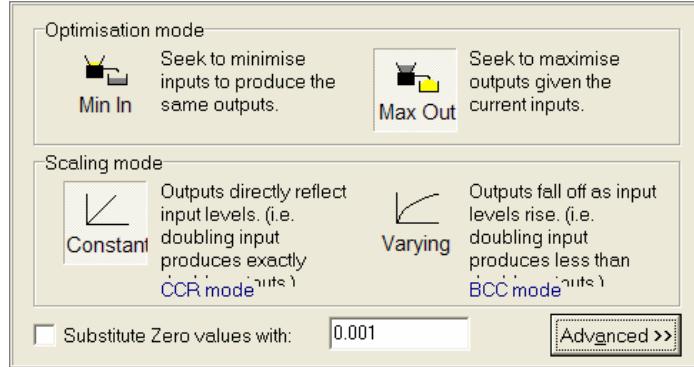


Figure 28: The DEA Options dialog.

Two different types of DEA models are reflected in the scaling mode selections in the above dialog. The more basic of the two is the CCR model, initially proposed by Charnes, Cooper, and Rhodes in 1978. A detailed explanation of the CCR model can be found in chapters 2 and 3 of (Cooper, 2000). The other alternative is the BCC model (Banker-Charnes-Cooper) model which is characterized by variable returns to scale (see Chapter 4 of (Cooper, 2000) for a detailed explanation of this model). The other consideration is in the optimization mode we choose. For this example, we will first seek to maximize outputs¹⁴ and assume constant returns to scale. In another example, we will use the BCC model which assumes variable returns to scale.

5.3.3. Analysis and Results Using the CCR Model.

When we run the analysis, the first results we see are the efficiency scores. Schools that receive a score of 100 are considered efficient. Those that receive a score of less than 100 are considered inefficient (the score is the percent of efficiency relative to one or more of the efficient school that make up the reference set for the inefficient school).

¹⁴ Referred to as an output-oriented model in (Cooper, 2000:58, 103).

Data Viewer

Efficiency Scores

Name	Score
CENTRAL MISSOURI STATE UNIV	100.00
UNIV OF HAWAII AT MANOA	100.00
UNIVERSITY OF GUAM	94.21
THE JOHNS HOPKINS UNIVERSITY	91.86
UNIV OF CALIFORNIA-DAVIS	90.41
CLARKSON UNIVERSITY	90.17

Figure 29: The Efficiency Scores table.

In this example, only Central Missouri State University and the University of Hawaii at Manoa are considered efficient. A graphical summary of efficiency scores is available, showing a histogram of schools that scored within a certain range.

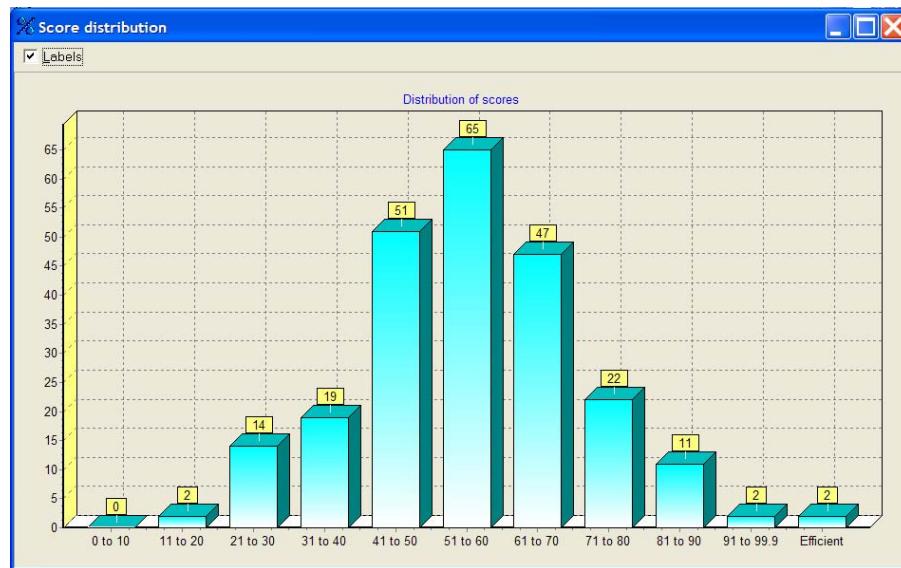


Figure 30: The Scores Distribution chart.

We can see here that the bulk of the schools (153 out of 235) in our dataset have efficiency scores between 41 and 70. Another way to see this distribution is by examining the frontier plot, which plots each school relative to the efficient frontier.

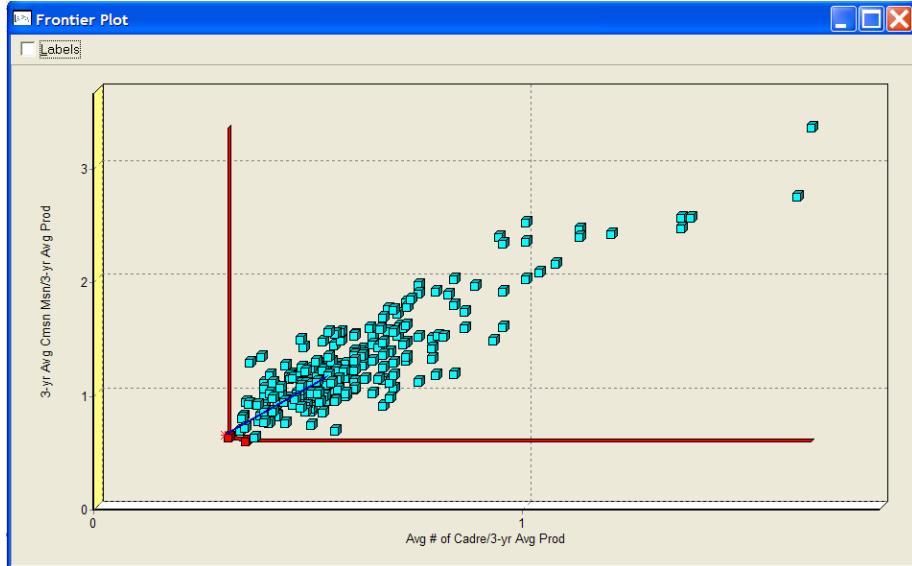


Figure 31: The frontier plot.

The efficient frontier always touches at least one point. In this case the two points on the frontier are the two school identified in the efficiency scores report at efficient. The region above and to the right of the efficient frontier (where all of our other data points reside) represents the production possibility set. Each point not on the frontier has its efficiency measures relative to a reference set that consists of at least one of the points on the efficient frontier (possibly both). In general, the further a school is from the efficient frontier, the lower its efficiency score. To see what other useful information can be gleaned from this analysis, we will examine a couple of school from the dataset more closely.

5.3.3.1. Example School 1: Appalachian State University.

Appalachian State University received an efficiency score of 56.9 percent. This places it firmly in the center of the distribution of efficiency scores. Its reference set is Central Missouri State University.

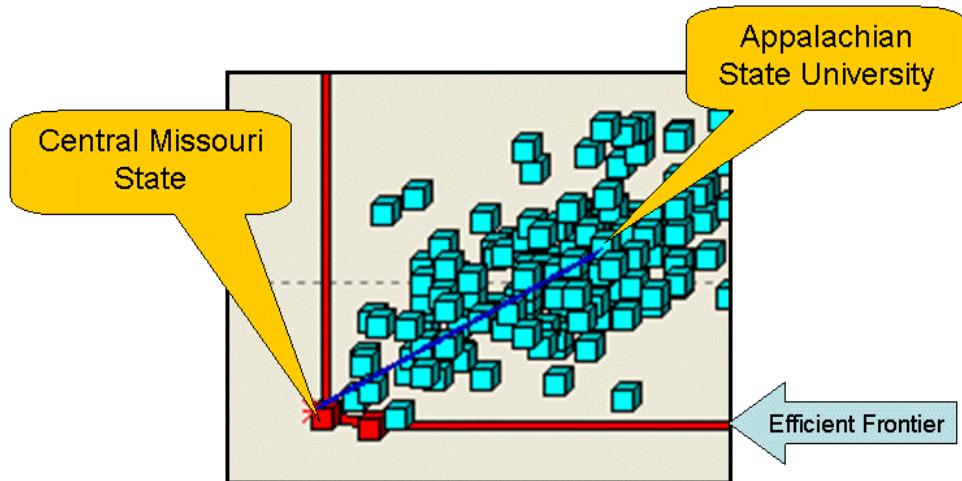


Figure 32: Comparison of Appalachian State to its reference set on the frontier plot.

Frontier Analyst also gives us additional tools, one of which allows us to examine potential areas of improvement. This information is provided by default in graphical form, but the numerical values can also be shown in a table (the entire graph window and part of the table window are shown in Figure 33 – the two different representations can be selected by using the buttons on the left side of the window).

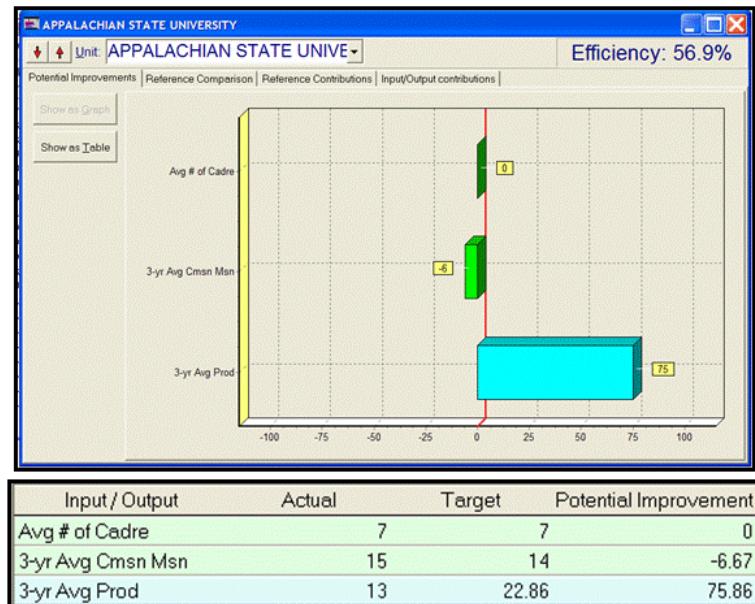


Figure 33: Potential improvements report (graph and table) for Appalachian State.

Note that this is not the only possible way to improve the efficiency of Appalachian State. In fact, if the suggested improvements were made exactly as indicated (by achieving the targets

for each variable), it would place Appalachian State on the efficient frontier precisely where Central Missouri State is located (since Central Missouri State is the only school in its reference set). In truth, any adjustment that places Appalachian State University on the efficient frontier will make the school efficient and place it among the reference set. Specifically, it would need to meet one of the following:

$$\begin{aligned}
 1) \frac{\text{Avg\#of Cadre}}{3 - \text{Yr Avg Prod}} &= 0.306 \text{ for } \frac{3 - \text{Yr Avg Cmsn Msn}}{3 - \text{Yr Avg Prod}} \geq 0.612 \\
 2) \frac{3 - \text{Yr Avg Cmsn Msn}}{3 - \text{Yr Avg Prod}} &= 0.577 \text{ for } \frac{\text{Avg\#of Cadre}}{3 - \text{Yr Avg Prod}} \geq 0.346 \\
 3) \text{On the line } \frac{3 - \text{Yr Avg Cmsn Msn}}{3 - \text{Yr Avg Prod}} &= -0.035 \left(\frac{\text{Avg\#of Cadre}}{3 - \text{Yr Avg Prod}} \right) + 0.623 \\
 \text{for } 0.306 < \frac{\text{Avg\#of Cadre}}{3 - \text{Yr Avg Prod}} &< 0.346
 \end{aligned}$$

Equation 3: Mathematical description of the efficient frontier.

Of course, any adjustment that causes Appalachian State University to outside of the production possibility set would cause it to replace one or both of the schools in the current reference set, assuming there are no changes elsewhere. Frontier Analyst also provides a reference comparison:

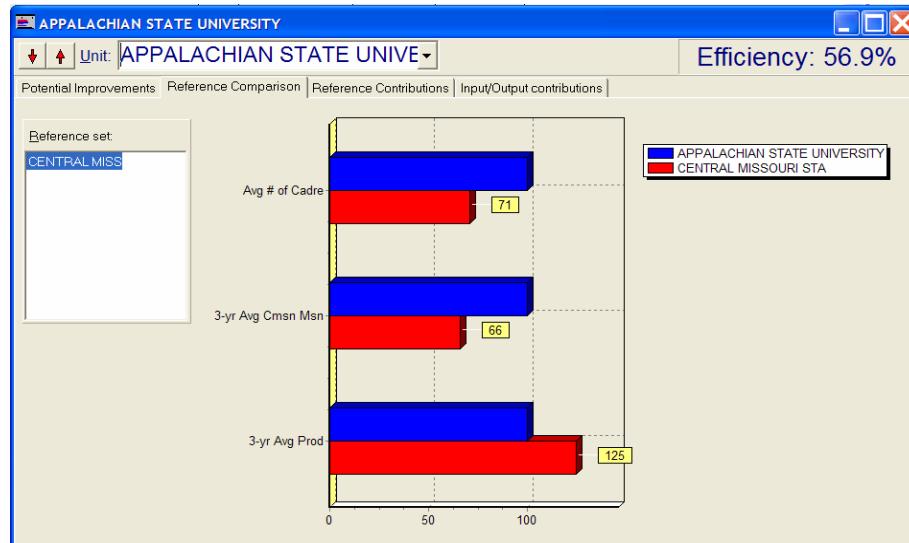


Figure 34: By-variable comparison to a reference set DMU.

This graph shows a direct comparison for each variable between Appalachian State University and Central Missouri State University. The numbered labels to the right of each bar for the reference indicate a percentage comparison to the school being examined. In this case, it shows that Central Missouri State has 71% of the average cadre and 66% of the average mission that Appalachian State did for the same period, but produced 125% of Appalachian State's average production.

5.3.3.2. Example School 2: University of Portland.

The University of Portland received an efficiency score of 81.4%, placing it in the upper tail of the distribution of efficiency scores. It has both efficient schools in its reference set.

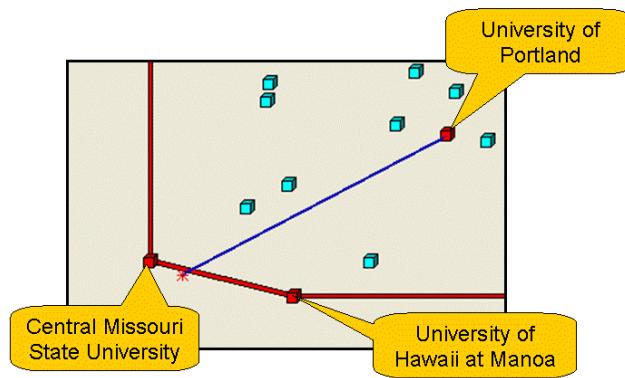


Figure 35: Comparison of the University of Portland to its reference set on the frontier plot.

The only potential improvement that Frontier Analyst recommends is to increase production by 22%. This adjustment would place the point in the frontier plot at the other end of the blue line segment shown in the above plot.

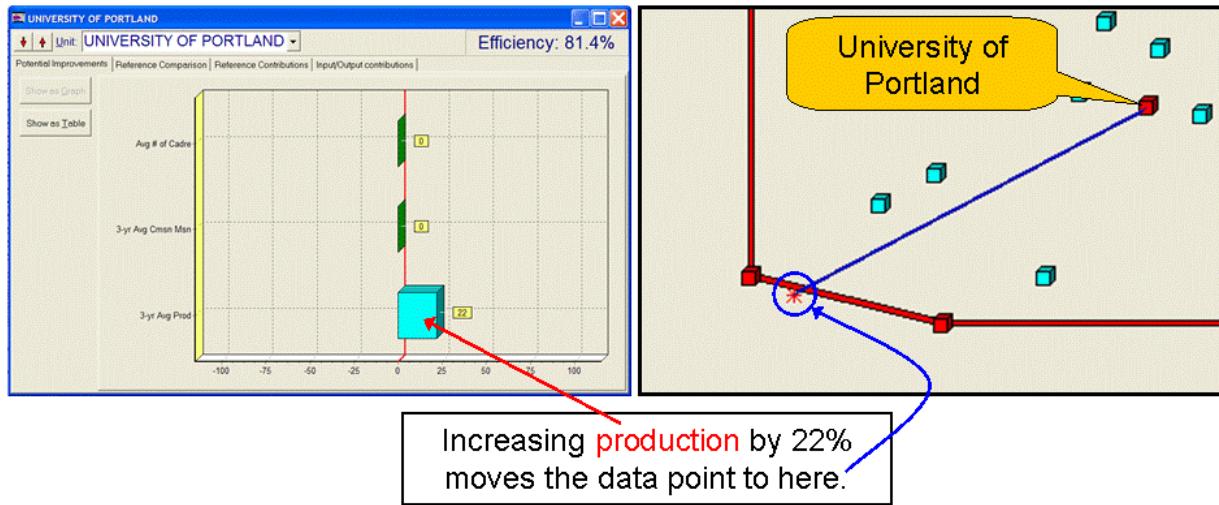


Figure 36: The effect of implementing the potential improvements.

When we look at the reference comparison, there are now two graphs available, since there are two schools in the reference set. Each comparison shows the relative percentages of the respective reference school when compared with the school of interest. In both graphs, the University of Portland is represented in blue.

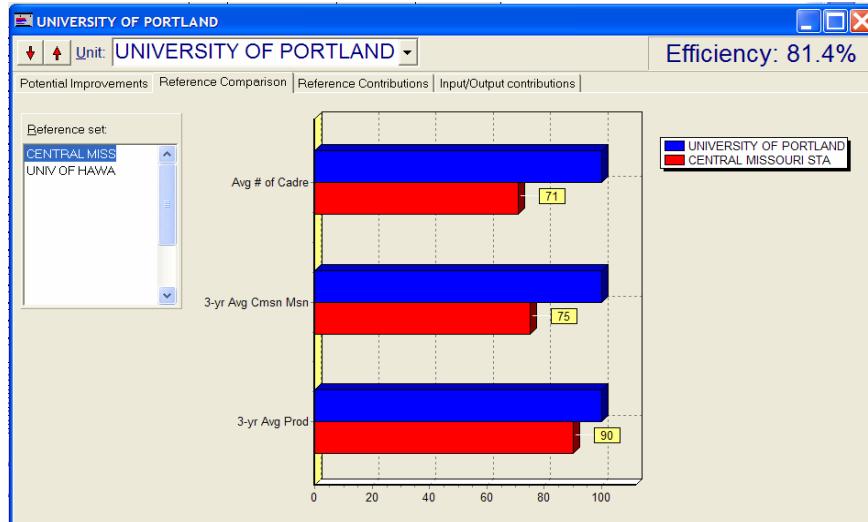


Figure 37: By-variable comparison of the University of Portland to Central Missouri State University.

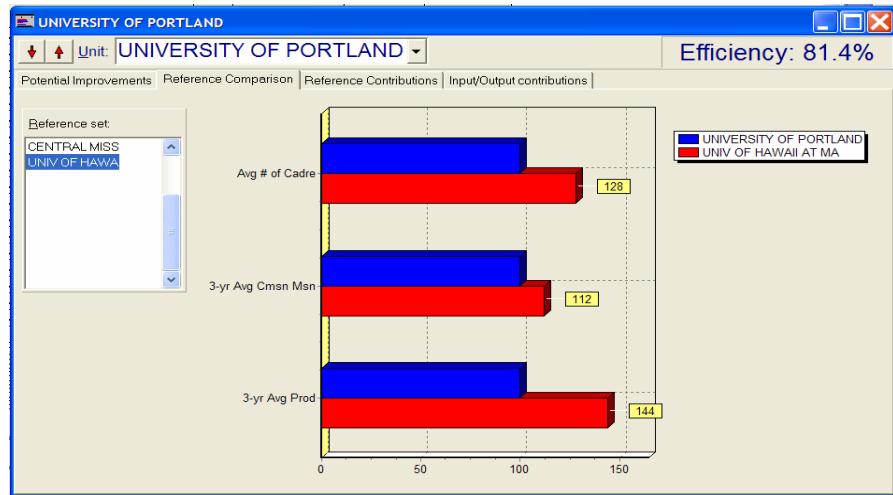


Figure 38: By-variable comparison of the University of Portland to the University of Hawaii at Manoa.

These graphs demonstrate a scale difference between the two reference schools, which is revealed when compared with the University of Portland. The University of Hawaii at Manoa, compared in the graph at bottom, has larger values for every variable compared to the University of Portland, while Central Missouri State has smaller values for every variable. In both cases the schools in the reference set show their highest comparison percentages in the area of production.

5.3.4. Analysis and Results Using the BCC Model.

Under the CCR model, constant returns to scale is assumed. This essentially means that a doubling of all inputs leads to a doubling of all outputs. If the DMUs do not exhibit this behavior, then they are exhibiting variable returns to scale. Because of this, when we use the BCC model, DMUs are compared in terms of efficiency to other DMUs of the same or similar scale as opposed to every DMU in the data set. DMUs will be reported to be at least as efficient as they were under the CCR model, and some DMUs will be reported as more efficient. To run this analysis, the only item that needs to be changed is the scaling mode in the analysis options:

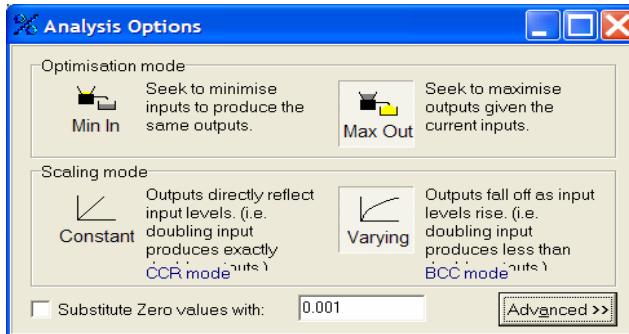


Figure 39: Changing to varying returns to scale (BCC model) in the DEA options dialog.

We now select varying returns to scale, which will then utilize the BCC model. We will maximize outputs, as before. After running the analysis, we again see the efficiency scores as the first output:

Efficiency Scores									
Details	Edit data	Show all	100%	<100 %	Unsort	A _Z ↓	9 ₁ ↓	1 ₉ ↓	Summary
235 units.									
Name	Score	Scale							
EDINBORO UNIV OF PENNSYLVANI	100.00	constant							
GEORGE MASON UNIVERSITY	100.00	constant							
SOUTHERN ILLUNIV-EDWARDSVIL	100.00	constant							
AUGUSTA STATE UNIVERSITY	100.00	constant							
CENTRAL MISSOURI STATE UNIV	100.00	constant							
UNIV OF HAWAII AT MANOA	100.00	constant							
UNIVERSITY OF PR MAYAGUEZ	100.00	constant							
UNIVERSITY OF GUAM	99.23	decreasing							
UNIV OF CALIFORNIA-DAVIS	96.60	increasing							
THE JOHNS HOPKINS UNIVERSITY	95.42	increasing							
CLARKSON UNIVERSITY	93.67	increasing							
THE PENNSYLVANIA STATE UNIV	91.97	increasing							
UNIVERSITY OF NOTRE DAME	91.92	increasing							
TEXAS STATE UNIV-SAN MARCOS	91.87	increasing							
UNIVERSITY OF CENTRAL FLORIDA	90.12	increasing							
BRIGHAM YOUNG UNIVERSITY	89.44	increasing							
CAMPBELL UNIV	88.52	increasing							
XAVIER UNIVERSITY	87.25	increasing							
NIAGARA UNIVERSITY	85.42	decreasing							
AUSTIN PEAY STATE UNIVERSITY	84.82	increasing							
WASHINGTON UNIVERSITY	84.56	decreasing							
UNIVERSITY OF PORTLAND	84.53	increasing							

Figure 40: Efficiency Scores table from the BCC model.

When the BCC model is used, an additional column appears labeled “Scale”. This indicates the type of returns to scale that a particular school exhibits. The University of Guam shows a decreasing return to scale, which means than a doubling of all inputs leads to something less than a doubling of outputs. Likewise, the University of California at Davis shows an increasing return to scale, which means that a doubling of all inputs leads to something more than a doubling of outputs. Note that there are several more schools that are identified as ‘relatively efficient’ than we had under the CCR model. This is a typical result of changing to a

varying returns-to-scale model from a constant returns-to-scale model. Overall, however, the distribution of efficiency scores has not changed much from the CCR model (compare Figure 41 to Figure 30).

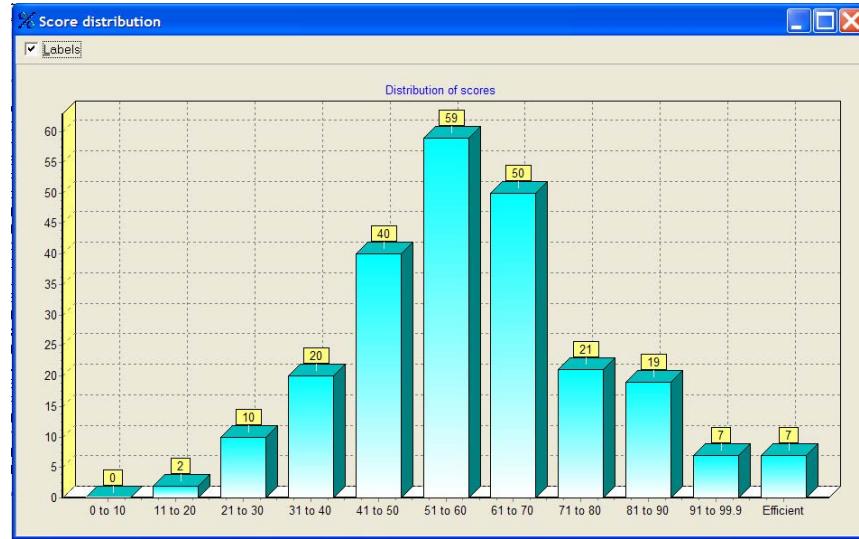


Figure 41: Distribution of Scores chart from the BCC model.

5.3.5. A Broader View at Future Application.

In actual use, there are more inputs that have an effect on production than the two included in our example. Several of these were listed in the introduction to this chapter. In general, it is possible to “inflate” the number of DMUs that will be identified as efficient, but normally this becomes a problem when the number of DMUs is less than the total of the input and output variables (Cooper, 2000:103). This would not be a problem if we are examining all of the schools at once, but should be considered if the data is examined in subsets. One way to reduce the number of variables in the analysis is to remove inputs that are strongly correlated with another input. Frontier Analyst has a tool to assist with this – once the analysis has been run, you can create a scatterplot (Frontier calls it an x-y plot) of any two variables in the data set. The window includes a correlation coefficient for the two variables being examined (shown in the upper-right corner of Figure 42).

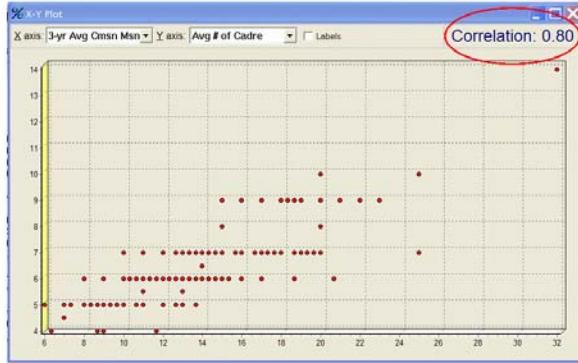


Figure 42: An x-y plot of two input variables.

A detailed analysis of USAREC recruiting practices, along with the development of a production function that incorporates DEA is given in (Brence, 2004). The regression model explained in section 3.2.1 of that paper is roughly analogous to the current STC model, although the variables used to determine the recruitable market are necessarily different (as are the time frames of interest). The addition of DEA into the model was done by analyzing the data of several inputs relevant to recruiting and recruiter allocation. If USACC wishes to examine alternate models (with possibly greater fidelity) once the OCMPS is complete, an approach similar to the one taken in (Brence, 2004) could be explored.

Chapter 6: Conclusion

The STC model uses an approach that makes sense with respect to USACC's current business practices. Its philosophy is consistent with the need to assign mission numbers to ROTC detachments in accordance with recruitable markets assessments. However, there is still enough variability in the recruitable market estimates to keep the model in its current role (a starting point that can be negotiated), as opposed to one with enough fidelity to eliminate the need for extensive mission adjustments. It is possible that, as the OCMPS proceeds, this variability will be reduced, or new and better indicators of recruitable market size may present themselves. Until that time, we have geared our recommendations toward three short-term goals. First, simplify the implementation of the model to make it leaner and easier to work with. Second, utilize built-in Excel functions wherever possible to reduce or eliminate mathematical errors, incorrect value assignments, and data duplication. Finally, utilize historical production data to assign different minimum mission values, where appropriate, in order to assist in meeting the increased overall mission. Beyond the near term, we also recommend applying the DEA approach to available data in order to separate efficient and inefficient performers. This will provide additional data on which to base resource allocation decisions.

Appendix A: An Explanation of the Ryan-Joiner Test for Normality

A normal probability plot is often used to determine whether it is plausible that a certain set of data was sampled from a normal population. This is done by ordering the sample x_i 's from smallest to largest, $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$, and then plotting the points $(x_{(i)}, y_i)$, where $y_i = \Phi^{-1}((i - 0.5)/n)$. The closer this plot resembles a straight line, the more plausible it is that the sample was taken from a normal distribution. This method is easy to use, but can become cumbersome if you have to repeat it more than a few times. The Ryan-Joiner test for normality allows us to take a quantitative measure of the extent to which points cluster about a straight line. A hypothesis test is used to make a determination based on a chosen significance level. The details are explained here:

Let $y_i = \Phi^{-1}[(i - 0.375)/(n + 0.25)]$ and compute the sample correlation coefficient r for the n pairs $(x_{(1)}, y_1), \dots, (x_{(n)}, y_n)$. The Ryan - Joiner test of

H_0 : the population distribution is normal

versus

H_A : the population distribution is not normal

consists of rejecting H_0 when $r \leq c_\alpha$ for significance level α .

A more detailed explanation can be found in Devore, page 651.

Appendix B: An Detailed Explanation of the Simulation Used to Determine the Alternate Minimum Mission Threshold

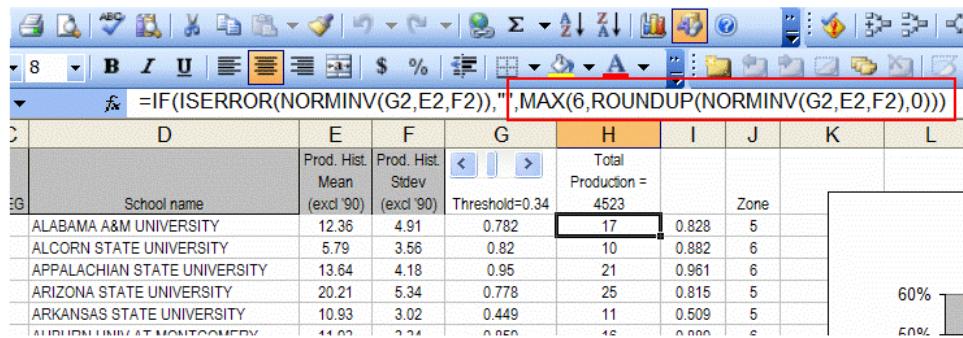
Doing Monte Carlo simulation in Microsoft Excel, where random values are simulated from a normal distribution, is a straightforward affair. You can generate these random values with a compound function that utilizes the **NORMINV** and **RAND** functions with the following statement:

$$=\text{NORMINV}(\text{RAND}(),0,1)$$

In this case, we have simulated a value from the standard normal distribution (see Microsoft Excel Help for details on the syntax of these functions). In order to simulate 500 values from this distribution, this function can be copied into 500 cells – the **RAND** function will produce a different pseudorandom probability each time it is called, which then gives a new simulated value from the distribution. For our purposes, we needed the simulation to only draw random sample values from a portion of the normal distribution used to approximate each school's production – specifically, only values that were greater than or equal to a specified percentile. To do this, we used the **RANDBETWEEN** function (instead of the **RAND** function) to generate the desired random probabilities.

C	D	E	F	G	H
EG	School name	Prod. Hist. Mean (excl '90)	Prod. Hist. Stdev (excl '90)		Total Production =
	ALABAMA A&M UNIVERSITY	12.36	4.91	0.782	17
	ALCORN STATE UNIVERSITY	5.79	3.56	0.82	10
	APPALACHIAN STATE UNIVERSITY	13.64	4.18	0.95	21
	ARIZONA STATE UNIVERSITY	20.21	5.34	0.778	25

In our simulation, the value stored in cell G275 is an integer between 1 and 999, and is controlled by the slider at the top of column G. The function is the same in each row and generates random probabilities between the threshold (G275/1000) and 0.999. This value is then used in determining the production values for the simulation – this is done with another compound expression that uses the MAX, ROUNDUP, and NORMINV functions:



The first part of the function uses the IF and ISERROR functions to deal with blank cells or missing data. The function that actually generates our random production values is outlined in red. The effect of this function is to generate production values that are always greater than a school's 34th percentile, and are never less than six. The values in column H are totaled to produce a value for overall production.

In order to determine the frequency of total production values that exceed 4400, we need to run this simulation several times. You can manually do this by pressing the F9 key (which will recalculate the entire worksheet), but it would be tedious to do this 500 times and write the results down after each run. There are specialized Excel add-ins that can handle this task (such as Palisade's @Risk for Excel[®]), but it can be done with a “trick” that uses Excel's table function. The steps are as follows:

1. Use the SUM function to calculate the total production of all schools. An example of this is shown below, using data from 10 schools (the function is stored in cell E12).

\bar{X}	s	Probability	Production
12.36	4.91	0.928	20
5.79	3.56	0.772	9
13.64	4.18	0.741	17
20.21	5.34	0.41	19
10.93	3.02	0.72	13
11.93	3.34	0.912	17
17.50	6.12	0.482	18
7.93	2.40	0.731	10
13.71	4.66	0.611	16
TOTAL			139

2. Make a reference to the summation cell in another cell on the worksheet (i.e enter “=E12) in another cell.

Probability	Production	=E12
0.928	20	
0.772	9	
0.741	17	
0.41	19	
0.72	13	
0.912	17	
0.482	18	
0.731	10	
0.611	16	
TOTAL		139

3. Starting one cell below and to the right of the last cell, start a sequence of integers beginning with 1 and ending with the highest number of simulation runs you want to execute (we'll use 10).

Probability	Production	141
0.775	17	1
0.647	8	2
0.926	20	3
0.652	23	4
0.871	15	5
0.728	14	6
0.351	16	7
0.902	12	8
0.655	16	9
TOTAL		10

4. Highlight the block of cells as illustrated below.

D	E	F	G	H
Probability	Production			141
0.775	17		1	
0.647	8		2	
0.926	20		3	
0.652	23		4	
0.871	15		5	
0.728	14		6	
0.351	16		7	
0.902	12		8	
0.655	16		9	
TOTAL	141		10	

5. Select “Table..” from the Data menu. The Table dialog appears as shown below.

C	D	E	F	G	H
s	Probability	Production			141
4.91	0.775	17		1	
3.56	0.647	8		2	
4.18	0.926	20		3	
5.34	0.652	23		4	
3.02	0.871	15		5	
3.34	0.728	14		6	
6.12	0.351	16		7	
2.40	0.902	12		8	
4.66	0.655	16		9	
TOTAL	141		10		

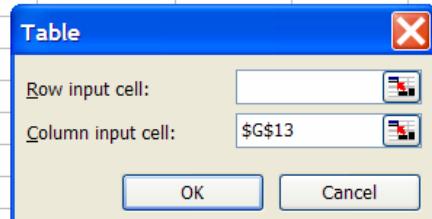
Table

Row input cell:

Column input cell:

6. Click in the “Column input cell”, then click in an empty cell (one you know you will not use later). We’ll use the cell just below the sequence for this example.

D	E	F	G	H
Probability	Production			
0.775	17			
0.647	8			
0.926	20			
0.652	23			
0.871	15			
0.728	14			
0.351	16			
0.902	12			
0.655	16			
TOTAL	141			
			141	
			1	
			2	
			3	
			4	
			5	
			6	
			7	
			8	
			9	
			10	



7. Click “OK”. The simulation will start (if you have AutoCalculate turned off, you will have to press the F9 key). The cells to the right of the sequence will populate with results from individual simulation runs. This process can be slow – in this example, we only ran 10 simulation runs, and Excel populated the cells in about three seconds (on a 2GHz Pentium M). In the actual simulation (which we ran 500 times), run time was about 90 seconds. You will get results similar to this:

A	B	C	D	E	F	G	H
1							
2	\bar{x}	s	Probability	Production		134	
3	12.36	4.91	0.958	21		1	148
4	5.79	3.56	0.492	6		2	161
5	13.64	4.18	0.367	13		3	142
6	20.21	5.34	0.671	23		4	145
7	10.93	3.02	0.613	12		5	149
8	11.93	3.34	0.494	12		6	132
9	17.50	6.12	0.403	16		7	148
10	7.93	2.40	0.433	8		8	140
11	13.71	4.66	0.964	23		9	148
12			TOTAL	134		10	138
13							

8. In our simulation, we were interested in the proportion of simulation runs that resulted in a total production of at least 4400. This can be done using the COUNTIF function. In this example, we will look for simulation runs that resulted in a total of at least 140. You can do this with the following expression:

The screenshot shows a Microsoft Excel spreadsheet with data in columns A through H. The formula bar at the top displays the formula `=COUNTIF(H3:H12,">=140")`. A red arrow points from the formula bar to the range H3:H12, which is highlighted with a red border. The data in column H consists of 10 numerical values: 148, 161, 142, 145, 149, 132, 148, 140, 148, and 138. The cell H14 contains the result of the COUNTIF function, which is 8. A blue callout bubble with the text "Function is in this cell" points to the cell H14.

	A	B	C	D	E	F	G	H
1								
2		\bar{x}	s	Probability	Production			
3		12.36	4.91	0.82	17	1	148	
4		5.79	3.56	0.932	12	2	161	
5		13.64	4.18	0.934	20	3	142	
6		20.21	5.34	0.499	21	4	145	
7		10.93	3.02	0.965	17	5	149	
8		11.93	3.34	0.954	18	6	132	
9		17.50	6.12	0.843	24	7	148	
10		7.93	2.40	0.524	9	8	140	
11		13.71	4.66	0.831	19	9	148	
12		TOTAL		157	10	138		
13								
14								8
15								

To get the proportion of runs meeting the criteria, just divide the expression by the number of runs (i.e. `“=COUNTIF(H3:H12,”>=140”)/10`).

Appendix C: Table of Abbreviations

ADM	Academic discipline mix
BCC	Banker-Charnes-Cooper DEA model (variable returns to scale)
CCR	Charnes-Cooper-Rhodes DEA model (constant returns to scale)
DEA	Data Envelopment Analysis
DMU	Decision-making unit
HBCU	Historically black colleges and universities
HIST	Historical Data Module
MDPM	Missions Data Processing Module
MJC	Military Junior College
NCES	National Center for Education Statistics
OAS	Officer Accession Strategy
OCMPS	On-Campus Market Potential Study
ODM	OCPMS Data Module
ORCEN	Operations Research Center of Excellence
ROTC	Reserve Officer Training Corps
SAL	Scholar-Athlete-Leader
SDM	School Data Module
SMC	Senior Military College
STC	Shaping the Cohort
SUM	Data Summary Module
USACC	United States Army Cadet Command
USAREC	United States Army Recruiting Command
YATS	Youth Attitude Tracking Study

Bibliography

1995 Annual Defense Report, <http://www.defenselink.mil/execsec/adr95/toc.html>

Brence, J., Kwinn, M., and Thomas, D., "Qualitative and Quantitative Analysis for U.S. Army Recruiting Input Allocation", presented at the 72nd Military Operations Research Society Symposium, Naval Postgraduate School, Monterey, CA, June 2004.

"Brief History of Army ROTC", Princeton Review,
<http://www.princetonreview.com/cte/articles/military/rotchist.asp>

Cooper, W., Seiford, L., and Tone, K., "Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software", Kluwer Academic Publishers, Boston, 2000.

Devore, Jay L., "Probability and Statistics for Engineering and the Sciences", Sixth Edition, Thomson Brooks/Cole, 2004.

Hoff, Michael, Director, Recruiting & Retention Operations Directorate (RROD), USACC.
Personal interview. Fort Monroe, VA, 23 November 2004.

Orvis, B. and Asch, B., "Military Recruiting: Trends, Outlooks, and Implications", RAND, 2001.

Warner, William and other USACC personnel. Conference. Fort Monroe, VA, 11 March 2005.

Warner, William, Division Chief, Operations Analysis Division, RROD, USACC, "Shaping the Cohort – Model & Process", briefing given to the author, Fort Monroe, VA, 14 September 2004.

Distribution List

The list indicates the complete mailing address of the individuals and organizations receiving copies of the report and the number of copies received. Due to the Privacy Act, only use business addresses; no personal home addresses. Distribution lists provide a permanent record of initial distribution. The distribution information will include the following entries:

NAME/AGENCY	ADDRESS	COPIES
Author(s)	Department of Systems Engineering Mahan Hall West Point, NY 10996	2
Client(s)	US Army Accessions Command 1307 Third Avenue Ft. Knox, KY 40121	1 each
	US Army Cadet Command Recruiting, Retention & Operations Directorate (RROD) Ft. Monroe, VA 236651	
Dean, USMA	Office of the Dean Building 600 West Point, NY 10996	1
Defense Technical Information Center (DTIC)	ATTN: DTIC-O Defense Technical Information Center 8725 John J. Kingman Rd, Suite 0944 Fort Belvoir, VA 22060-6218	1
Department Head-DSE	Department of Systems Engineering Mahan Hall West Point, NY 10996	1
ORCEN	Department of Systems Engineering Mahan Hall West Point, NY 10996	5
ORCEN Director	Department of Systems Engineering Mahan Hall West Point, NY 10996	1
USMA Library	USMA Library Bldg 757 West Point, NY 10996	1

REPORT DOCUMENTATION PAGE – SF298

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 13-07-2005	2. REPORT TYPE Technical Report	3. DATES COVERED (From - To) 07 Sep 04-13 Jul 05		
4. TITLE AND SUBTITLE Shaping the ROTC Cohort		5a. CONTRACT NUMBER		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Major Wiley P. Rittenhouse Lieutenant Colonel Michael J. Kwinn, Jr.		5d. PROJECT NUMBER DSE-R-0505		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Operations Research Center of Excellence Dept. of Systems Engineering US Military Academy West Point, NY 10996		8. PERFORMING ORGANIZATION REPORT NUMBER DSE-TR-0505		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Army Accessions Command 1307 Third Avenue Fort Knox, KY 40121 (502) 626-0308		10. SPONSOR/MONITOR'S ACRONYM(S) USAAC, USACC		
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION / AVAILABILITY STATEMENT Distribution A: Approved for public release; distribution is unlimited.				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT U.S. Army Cadet Command (USACC) has developed a new commissioning model - Shaping the Cohort (STC) - to meet the future needs of the Army for commissioned officers. It is designed to shape each cohort to meet the Army's specific needs in terms of component, academic disciplines, race/ethnic makeup goals, gender, and targeted missions. STC does this by determining and examining the "prime market" at a university and basing the detachment's mission on penetration of that market as opposed to one based on past performance. It is believed that the STC model improves the method of determining missions. To determine market potential, USACC conducted two surveys that included 62 colleges and universities and over 7600 students. The goals of the survey were to determine knowledge and perception of Army ROTC among students, segmentation of local markets, how the school markets differed, and the characteristics that could lead to participation in Army ROTC. The data gathered is used to determine how many students at each school fit the criteria for the prime market. The Operations Research Center of Excellence (ORCEN) at West Point has provided an independent assessment of the model's adequacy and to determine if and how it can be improved. We begin this report with a needs analysis to determine the real problem the model attempts to address. We then present our analysis of the STC model and the process for determining missions. In the next section we identify recommendations to improve the model and conclude with suggested additional tools for conducting this analysis.				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON MAJ Wiley Rittenhouse
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	None	55
		19b. TELEPHONE NUMBER (include area code) 845-938-5168		